

Levinneisyysmallinnus uhanalaistuvien päiväpetolintujen esiintymisen kartoittamisessa Suomessa

Petteri Kiilunen

Pro gradu -tutkielma
Maantieteen tutkimusyksikkö
Oulun yliopisto
2.12.2020



Yksikkö: Maantieteen tutkimusyksikkö		Pääaine: Maantiede	
Tekijä (Sukunimi ja etunimet, myös entinen sukunimi): Kiilunen Petteri Juho-Matti		Opiskelija-numero: 2370385	Tutkielman sivumäärä: 82 + V liit.
Tutkielman nimi (suomeksi; muun kielinen nimi ilmoitetaan vain jos se on tutkielman kieli): Levinneisyysmallinnus uhanalaistuvien päiväpetolintujen esiintymisen kartoittamisessa Suomessa			
Asiasanat: Levinneisyys, levinneisyysmallinnus, päiväpetolinnut, ympäristömuuttajat, pedot, habitaatti, suojele			
Tiivistelmä (kirjoitetaan vapaamuotoisesti, selväsanaisesti ja lyhyin lauserakentein, ks. ohje seuraavalla sivulla):			
<p>Yhä useammat eliölajit ovat uhanalaistuneet viimeisten vuosikymmenten aikana globaalimuutoksen myötä. Lajien levinneisyyden tutkiminen voi paljastaa syitä uhanalaistumiselle ja auttaa tunnistamaan ja suojelemaan lajien kannalta tärkeitä elinympäristöjä. Eliöiden levinneisyysalue muodostuu monimutkaisten prosessien seurauksena, mutta elinympäristön eli habitaatin laadulla tiedetään olevan suuri vaikutus lajien säilymisessä. Pedot säätelevät eliöyhteisöjen tasapainoa ja ovat herkkiä esimerkiksi saalislajeissa ja habitaatissa tapahtuville muutoksille.</p> <p>Tutkielman tavoitteena oli tarkastella kahden päiväpetolintulajin, piekanan (<i>Buteo lagopus</i>) ja mehiläishaukan (<i>Pernis apivorus</i>), esiintymistä Suomessa hyödyntäen levinneisyysmallinnusta (<i>species distribution modeling</i>, SDM). Lajit eroavat toisistaan levinneisyydeltään, ravintotottumuksiltaan ja habitaatinvalinnaltaan. Levinneisyysmallinnus perustui vuosien 2000 ja 2015 välillä tehtyihin lajien esiintymishavaintoihin, jotka oli paikannettu koko Suomen kattaviin 10 km x 10 km tutkimusruutuihin. Levinneisyyttä selittävinä tekijöinä käytettiin lämpötilan ja sadannan keskiarvoja lajien kevätmuuton ja pesinnän kannalta tärkeinä ajankohtina sekä lajien habitaattien määrää ja saatavuutta kuvaavia maanpeite- ja maankäyttömuuttujia. Levinneisyysmallinnuksen avulla pyrittiin ennustamaan piekanan ja mehiläishaukan esiintymistä Suomessa ja selvittämään lajien kannalta suotuisimmat habitaatit. Mallien luotettavuutta ja ennustuskykyä tarkasteltiin monipuolisesti, jotta käytetyn aineiston käyttökelpoisuutta päiväpetolintujen levinneisyysmallinnukselle voitiin arvioida.</p> <p>Tulosten perusteella päiväpetolintulajien levinneisyyttä parhaiten ennustavat menetelmät olivat <i>generalized linear model</i> (GLM), <i>generalized additive model</i> (GAM), <i>generalized boosted model</i> (GBM) ja <i>random forest</i> (RF). Kyseiset menetelmät ovat hyvä lähtökohta valittaessa menetelmää päiväpetolintujen levinneisyyden mallintamiseen. Parhaiten päiväpetolintujen levinneisyyttä selitti lämpötila. Sadanta vaikutti negatiivisesti lajien levinneisyyteen. Maanpeitettä ja maankäyttöä kuvaavien muuttujien merkitys oli malleissa pienin ja ne näyttäisivät soveltuvan heikommin päiväpetolintujen levinneisyyden mallintamiseen. Levinneisyysmallien ennustuskyky oli hyvä ja mallit tuottivat uskottavia ennusteita lajien levinneisyydestä Suomessa. Mallien luotettavuuteen ja soveltuvuuteen tulee kuitenkin kiinnittää huomiota tutkimuskohtaisesti, jos tuloksia halutaan hyödyntää esimerkiksi jatkotutkimuksissa tai suojeletoimenpiteissä.</p>			
Muita tietoja:			
Päiväys: 04.12.2020 Oulu			

Sisällys

Johdanto	3
Levinneisyys eliömaantieteellisenä ilmiönä	6
Levinneisyysmallinnus tutkimusmenetelmänä	7
Levinneisyyteen vaikuttavat tekijät	10
Dispersaali	11
Abioottinen ympäristö	12
Bioottinen ympäristö	14
Päiväpetolinnut Suomessa – uhanalaistuvien lajien määrä on kasvussa	16
Uhanalaistuminen ei rajoitu tiettyihin habitaatteihin – esimerkkilajeina piekana ja mehiläishaukka	17
Käsitteellisestä mallista kohti levinneisyysmallinnusta	19
Tavoitteena ennustaa lajin potentiaalista tai toteutunutta levinneisyysaluetta	19
Levinneisyysmallinnus pohjautuu lajin yksilöiden esiintymiseen	22
Ympäristömuuttujat kuvaavat habitaatin ominaisuuksia	26
Tutkimusalue	31
Aineisto ja menetelmät	34
Tutkimusaineisto	34
Tutkimusvaiheet ja -menetelmät	36
Levinneisyysmallinnuksen menetelmien ennustuskyykyä tulee tarkastella aineistokohtaisesti	37
Levinneisyysmallinnuksen ennusteita tulee tarkastella aineistoon sopivilla arviointikriteereillä	41
Tulokset	44
Menetelmien ennustuskyyvyn tarkastelu ja vertailu	44
Mallien sovittaminen <i>Generalized Boosted Model</i> -menetelmällä ja luotettavuuden arviointi	49
Pohdinta	56
Tarkasteltujen menetelmien ennustuskyyvyssä ei havaittu suuria eroja	56
Ympäristömuuttujien tarkastelu voi paljastaa heikot levinneisyysmallit	57
Havaintojen ja muuttujien suhde vaikuttaa mallien luotettavuuteen	63
Ylisovittaminen, tausta-aineisto ja spatiaalinen autokorrelaatio vaikuttavat ennusteisiin	65
Aineiston rajoitteet eivät estäneet lajien levinneisyysmallinnusta	68
Päiväpetolintujen levinneisyyttä voidaan ennustaa useilla menetelmillä	70
Käytettyä aineistoa voidaan tukea ja hyödyntää monin eri tavoin	71
Johtopäätökset	73
Kirjallisuus	74
Liitteet	83

Johdanto

Jokaisella maailman mantereella on viime vuosisatojen ja -tuhansien saatossa tapahtunut lajiston ja ekosysteemien muutoksia (Lomolino ym. 2017). Muutoksia on dokumentoitu viimeisten parin vuosikymmenen aikana yhä enemmän ja enemmän niin ekosysteemien lajimäärissä, lajien yksilöiden määrässä kuin myös lajien levinneisyydessä (Root ym. 2003). Muutokset eivät ole kuitenkaan kohdistuneet tiettyihin eliöryhmiin tai elinympäristöihin eli habitaatteihin, vaan niitä havaitaan ympäri maapalloa lukuisissa lajeissa kasveista eläimiin (Walther ym. 2002), mikä on pakottanut tutkijat selvittämään ja ymmärtämään lajeihin vaikuttavia ympäristötekijöitä ja syitä uhanalaistumisen taustalla. Havaintojen laajuudesta ja muutosten nopeudesta johtuen tärkeimpänä taustasyynä pidetään globaalimuutosta (Chen ym. 2011). Globaalimuutos pitää sisällään useita, ympäri maailmaa vaikuttavia huolestuttavia ilmiöitä, kuten ilmastomuutoksen, luonnonvarojen kasvavan kulutuksen ja maankäytön muutokset, jotka yhdessä muuttavat lajien habitaatteja (Parmesan ym. 2006). Uhanalaistuneiden lajien määrä kasvaa vuosi vuodelta ja jopa aiemmin yleiset lajit ovat taantuneet nopeasti myös Suomessa (Laaksonen & Lehikoinen 2013).

Lajien levinneisyyteen vaikuttavien prosessien tunnistaminen on avainasemassa selvitetessä, miksi tietyt lajit reagoivat voimakkaammin habitaateissa tapahtuviin muutoksiin ja mitkä lajit ovat erityisen suuressa vaarassa kadota eri ekosysteemeistä (Brown ym. 2006; Lomolino ym. 2017). 2000-luvulla levinneisyyden tutkimuksen ja ekologisten menetelmien joukkoon on noussut vahvasti levinneisyysmallinnus (*species distribution modelling*, SDM), joka mahdollistaa lajien ja kokonaisten eliöyhteisöjen empiirisen tutkimuksen (Guisan & Zimmermann 2000; Guisan & Thuiller 2005; Elith & Leathwick 2009). Levinneisyysmallinnuksen avulla voidaan tutkia monipuolisesti erilaisia eliöiden levinneisyyteen sidoksissa olevia ilmiöitä (Franklin 2009; Peterson ym. 2011; Guisan ym. 2017) ja etsiä ratkaisuja ongelmiin, jotka vaativat eliöiden alueellisten esiintymisalueiden tuntemusta ja ymmärrystä levinneisyyteen vaikuttavista prosesseista (Austin 2002, 2007; Peterson ym. 2011). Levinneisyysmallinnuksen avulla voidaan myös esimerkiksi tutkia ja ennustaa ilmastomuutoksen ja habitaattikadon vaikutuksia eliöihin (Pearson & Dawson 2003; Thomas ym. 2004; Araújo ym. 2005; Wisz ym. 2008a), tarkastella vieraslajien leviämistä ja vaikutuksia uudessa ympäristössä (Gallien ym. 2012) ja ennusteita voidaan käyttää uhanalaisten lajien suojelutoimenpiteiden suunnittelussa (Kremen ym. 2008). Menetelmän kehittyminen tulee

tarjoamaan tulevaisuudessa myös lukemattomia uusia käyttötarkoituksia ja mahdollistamaan yhä monimutkaisempien tutkimusongelmien ratkaisun (Kissling ym. 2018).

Pohjoisten alueiden ekosysteemit ja yksittäiset lajit ovat haavoittuvaisia muutoksille esimerkiksi lämpötiloissa ja sadannassa tai suotuisten habitaattien määrässä (Parmesan ym. 2006). Ääriolosuhteisiin sopeutuneet lajit ovat myös epäsuotuisassa kilpailuasemassa uusille alueille levittäytyviin lajeihin verrattuna (Tylianakis ym. 2008; Chen ym. 2011; Virkkala & Rajasärkkä 2013). Yksi nopeasti uhanalaistuvista ryhmistä ovat päiväpetolinnut (*Accipitriformes*), jotka kärsivät muiden lajien tapaan muun muassa metsien muutoksista (Laaksonen & Lehikoinen 2013), perinnebiotooppien, kuten niittyjen ja katojen katoamisesta (Lehikoinen ym. 2019), ilmastonmuutoksesta (Schmidt ym. 2012; Terraube ym. 2015) ja talvehtimiseen liittyvistä häiriötekijöistä (Klaassen ym. 2014). Suomen 14 päiväpetolintulajista peräti yhdeksän on arvioitu uhanalaiseksi viimeisimmässä Suomen lajien uhanalaisuuden Punaisessa kirjassa (Lehikoinen 2019). Pedot vaikuttavat ekosysteemien tasapainoon esimerkiksi rajoittamalla hyödyntämiensä saalislajien yksilömääriä (Ritchie & Johnson 2009). Päiväpetolintujen on jopa todettu ylläpitävän korkeaa biodiversiteettiä eli luonnon monimuotoisuutta, jolloin niiden suojelu voi auttaa suojelemaan tehokkaasti konkaisia ekosysteemejä (Sergio ym. 2006).

Tutkielman tavoitteena on ennustaa uhanalaistuvien päiväpetolintujen levinneisyyttä Suomessa hyödyntäen levinneisyysmallinnuksen menetelmiä. Tutkimusaineisto koostuu kahden vaarantuneen päiväpetolintulajin, piekanan (*Buteo lagopus*) ja mehiläishaukan (*Pernis apivorus*) yksilöistä tehdyistä esiintymishavainnoista ja lajien habitaatteja kuvaavista ympäristömuuttujista, jotka esittävät alueellisia ilmastotekijöitä, maanpeitettä ja -käyttöä. Levinneisyysmallinnuksen ennusteet auttavat selvittämään lajeihin vaikuttavia ympäristötekijöitä ja rajaamaan lajeille suotuisat habitaatit Suomessa. Tavoitteena on myös tunnistaa parhaat menetelmät, joilla päiväpetolintujen levinneisyyttä voitaisiin mallintaa jatkossa vastaavan kaltaisella tutkimusaineistolla. Tutkimuskysymykset muodostuvat seuraavanlaisiksi:

- 1) Millaisilla levinneisyysmallinnuksen menetelmillä piekanan ja mehiläishaukan levinneisyyttä voidaan ennustaa Suomessa ja
- 2) voidaanko käytössä olevilla aineistoilla saada luotettavia tuloksia lajien esiintymisestä Suomessa?

Levinneisyys eliömaantieteellisenä ilmiönä

Lajien levinneisyys on eliömaantieteellisen tutkimuksen kulmakivi (Brown ym. 1996; Lomolino ym. 2017). Eliömaantieteelliselle tutkimukselle tyypillisiä tutkimuskohteita ovat esimerkiksi levinneisyysalueiden koko, rajat ja sijainti, levinneisyysalueita muokkaavat tekijät ja prosessit sekä levinneisyyden muutokset ajassa ja paikassa (Brown ym. 1996). Jotta voimme selittää, miksi laji esiintyy juuri siellä, missä sitä tavataan, tulee meillä olla kattava tieto lajin ominaisuuksista, vaatimuksista ja rajoitteista ja lajin asuttaman alueen ympäristöstä sekä siitä, millaiset prosessit lajiin ja ympäristöön vaikuttavat (Pulliam 2000). Jokaisella lajilla on yksilölliset tarpeensa, jotka ohjaavat lajin kykyä sietää erilaisia ympäristöjä ja sopeutua niihin, minkä vuoksi eri lajeja tavataan toisistaan poikkeavissa habitaateissa ja levinneisyysalueiden koko vaihtelee (Austin & Smith 1989).

Samalla maantieteellisellä alueella elävät lajin yksilöt muodostavat yhdessä oman ryhmän, paikallispopulaation. Lajin lukuisat paikallispopulaatiot muodostavat yhdessä lajin levinneisyyden eli jakautumisen maantieteelliseen tilaan. Tarkemman määrittelyn perusteella lajin levinneisyysalue kattaa sen maantieteellisen alueen, jolla esiintyy lisääntymiskelpoisia yksilöitä (Peterson ym. 2011: 24). Se, että lajin yksilöt kykenevät lisääntymään, voidaan katsoa olevan merkki alueen onnistuneesta kolonisaatiosta eli asuttamisesta: populaatio on tällöin sopeutunut ympäristön olosuhteisiin ja kykenee selviytymisen lisäksi kasvamaan kyseisessä habitaatissa (Pulliam 2000). Heikkolaatuisissa habitaateissa elävät populaatiot ovat alttiimpia esimerkiksi ympäristön muutoksille. Jotkin hyvän liikkumiskyvyn omaavat lajit, kuten muuttolinnut, taas saattavat esiintyä eri elinalueilla eri aikoihin, esimerkiksi vuodenkierron mukaan (Webster ym. 2002).

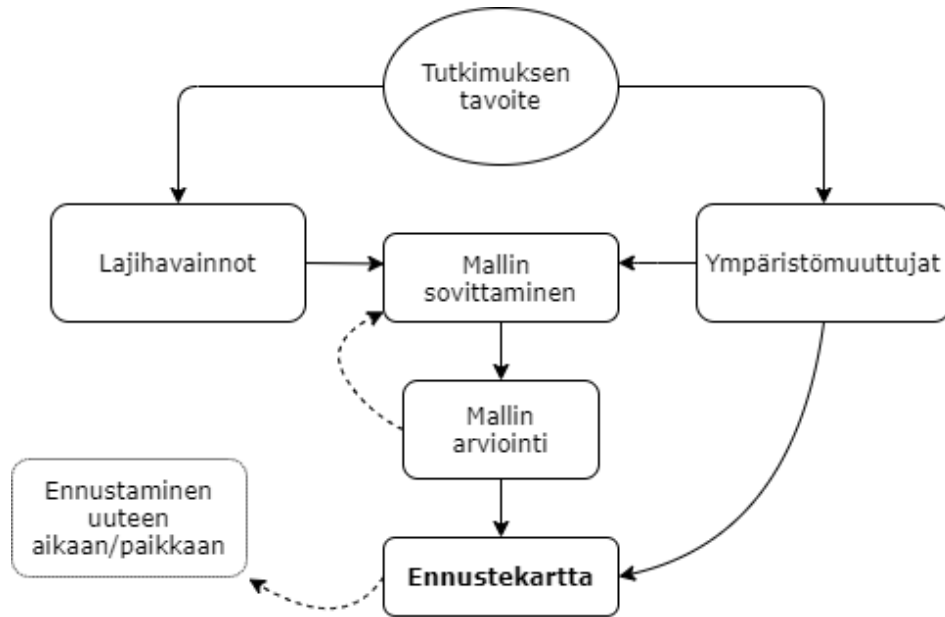
Levinneisyys ilmiönä on aina riippuvainen siitä spatiaalisesta mittakaavasta (*spatial scale*), esimerkiksi alueen koosta, jolla lajin esiintymistä kulloinkin tarkastellaan (Mackey & Lindenmayer 2001). Jos rajaamme tarkasteltavan alueen riittävän pieneksi, voimme merkitä ylös jokaisen lajin yksilön sijainnin ja määritellä kyseisen populaation levinneisyysalueelle rajat (Lomolino ym. 2017: 72–73). Tarkasteltavan alueen laajentaminen vaikuttaa havaitsemiemme yksilöiden ja populaatioiden määrään ja jakautumiseen ympäristössä muuttaen samalla levinneisyysalueen kokoa ja rakennetta. Levinneisyyteen kuuluu myös ajallinen näkökulma, sillä levinneisyysalue määritellään aina tiettyinä ajankohtana (Lomolino ym. 2017). Koska yksilöiden ja populaatioiden paikallinen ja alueellinen esiintyminen vaihtelee esimerkiksi yksilöiden liikkumisen myötä, lajin levinneisyys muuttuu

ajassa ja paikassa. Vaikka voisimme tiettyä hetkenä dokumentoida lajin jokaisen yksilön sijainnin, emme voi olla varmoja tilanteen pysyvyydestä tai siitä, millainen yksilöiden jakautuminen on ollut ennen kyseistä hetkeä (Gaston 2009).

Levinneisyyteen ja sen vaihteluun vaikuttaa voimakkaasti lajin sisäinen populaatiodynamiikka eli populaatioiden koossa ja rakenteessa tapahtuvat muutokset. Populaatioissa olevien yksilöiden määrä, populaatioiden luonnollinen ekstinktio eli häviäminen ja uusien populaatioiden syntyminen vaikuttavat levinneisyysalueen ja sen rajojen muodostumiseen (Soberón 2007). Tähän voivat johtaa esimerkiksi ympäristön tai lajin ympäristövaatimusten muuttuminen, ympäristössä tapahtuvat hetkelliset häiriötekijät ja populaatioiden luonnollinen stokastisuus eli satunnainen, esimerkiksi ympäristön vuotuisesta vaihtelusta johtuva vaihtelu populaation koossa ja kasvussa (Pulliam 2000; Holt ym. 2005). Pulliamin (1988) ajatus lähde- ja nielupopulaatioista yhdessä Hanskin (1999) metapopulaatio-teorian kanssa katsotaan tärkeäksi dynamiikaksi levinneisyysalueiden muodostumisen taustalla. Ympäristö vaikuttaa aina populaatioiden elinkelpoisuuteen, mutta itse vaikutuksen voimakkuus ja suunta riippuvat lajin omasta sietokyvystä ja kyvystä sopeutua vaihteleviin olosuhteisiin (Pulliam 2000; Soberón 2007).

Levinneisyysmallinnus tutkimusmenetelmänä

Yksi vakiintunut keino tutkia lajien levinneisyyttä ilmiönä on levinneisyysmallinnus, joka itse asiassa kattaa sisäänsä lukuisia erilaisia menetelmiä, joilla kaikilla on tavoitteena kuvailla ja mallintaa lajien ja habitaatin suhdetta sekä lajien levinneisyyttä paikassa ja ajassa (Guisan & Thuiller 2005; Franklin 2009; Guisan ym. 2017). Mallinnusmenetelmien avulla voidaan esimerkiksi ennustaa potentiaalista tai todellista lajin levinneisyysaluetta tai erilaisten habitaattien sopivuutta lajille (Elith & Leathwick 2009). Näitä ennusteita voidaan hyödyntää myös erilaisissa jatkotutkimuksissa (Pearson & Dawson 2003; Kremen ym. 2008). Levinneisyysmallinnukseen voidaan lukea kuuluvan kaikki ne menetelmät, jotka liittävät yhteen lajin yksilöistä tehdyt havainnot ja lajiin vaikuttavat ympäristötekijät tuottaen ennusteen, joka kuvaa lajin esiintymistodennäköisyyttä tai habitaattien sopivuutta tietyllä maantieteellisellä alueella (kuva 1, Franklin 2009; Guisan ym. 2017). Erilaisia mallinnustekniikoita hyödyntäen voidaan selvittää, miten ympäristötekijät vaikuttavat yksilöiden esiintymiseen ympäristössä ja ikään kuin luoda säännöt, joiden avulla tämä suhde voidaan siirtää maantieteelliseen tilaan kuvaamaan lajin levinneisyyttä (Austin 2002).



Kuva 1. Yksinkertainen kaavio levinneisyysmallinnuksen perusasetelmasta.

Levinneisyysmallinnuksen pohjana ovat **havainnot** tutkitun lajin yksilöiden esiintymisestä (Guisan & Zimmermann 2000). Havainnot voivat olla peräisin systemaattisista kartoituksista tai vaihtoehtoisesti satunnaisemmin kootuista lähteistä, kuten museonäytteistä, yksittäisten kansalaisten ja harrastajien ulkoisiin tietokantoihin tallentamista havainnoista (Kissling ym. 2018). Lajin yksilöiden havaittua jakautumista maantieteelliseen tilaan pyritään selittämään ympäristötekijöiden vaikutusta kuvaavilla **ympäristömuuttujilla** (Guisan & Zimmermann 2000), jotka kuvaavat lajin esiintymiseen ja runsauteen vaikuttavia todellisia ympäristötekijöitä (Austin 2002). Ympäristömuuttujat johdetaan tyypillisesti perinteisistä alueellisista ilmiöistä, kuten ilmastosta, topografiasta tai muista elottoman luonnon ympäristötekijöistä (Franklin 2009; Guisan ym. 2017). Levinneisyysmallinnuksessa voidaan tutkia yksittäisen lajin sijaan myös esimerkiksi eliöyhteisöjä kokonaisuutena tai alueellista biodiversiteettiä (Guisan & Thuiller 2005).

Lajihavaintojen ja ympäristömuuttujien yhdistäminen vaatii menetelmän, joka kykenee havaitsemaan niiden väliseen suhteeseen liittyvän säännönmukaisuuden ja sovittamaan sen malliksi (Guisan & Zimmermann 2000). Modernissa levinneisyysmallinnuksessa on hyödynnetty kymmeniä erilaisia menetelmiä, joista pääosa perustuu jossain muodossa tilastotieteen menetelmiin (Elith ym. 2006; Merow ym. 2014). Perinteisesti ekologisissa ja maantieteellisissä tutkimuksissa hyödynnetyt regressiomenetelmät ja niiden johdannaiset, kuten yleistetyt lineaariset mallit (GLM) ja yleistetyt additiiviset mallit (GAM), on todettu

levinneisyysmallinnuksessa suorituskyyisiksi ja luotettaviksi erityisesti tilanteissa, joissa on käytettävänä hyvälaatuista havaintoaineistoa lajin esiintymisistä ja poissaoloista (Elith & Leathwick 2009). Monimutkaisemmat, aineistoon nojaavat mallinnusmenetelmät, kuten tietokoneoppimisen menetelmistä johdetut mallit, kasvattavat suosiotaan johtuen menetelmien hyvästä suorituskyyvystä ja joustavuudesta, joka tapahtuu kuitenkin toisinaan mallien yleistämisen kustannuksella (Phillips ym. 2006; Elith & Leathwick 2009; Merow ym. 2014).

Menetelmän valintaa ohjaavat käytettävä aineisto ja tutkimuksen tavoitteet (Austin 2007; Franklin 2009; Guisan ym. 2017). Menetelmät eroavat toisistaan aineistovaatimuksiltaan ja kyyvyltään sietää ja mallintaa aineistossa esiintyvää vaihtelua, minkä vuoksi toisessa tutkimuksessa tehokkaimmaksi havaittu menetelmä ei välttämättä ole paras vaihtoehto omalle tutkimukselle, jonka aineisto ja tavoite ovat siitä poikkeavia (Elith & Leathwick 2009). Parasta, jokaiselle lajille ja jokaiseen ympäristöön sopivaa menetelmää ei ole voitu määritellä (Elith & Leathwick 2009; Qiao ym. 2015): menetelmien suorituskyyky ja mallien luotettavuus riippuu pitkälti havaintojen ja ympäristömuuttujien laadusta ja luotettavuudesta sekä tutkimuksessa käytetystä spatiaalisesta mittakaavasta (Araújo & Guisan 2006; Elith ym. 2006). Hyvä lähtökohta on käyttää vaihtoehtoisia menetelmiä mallien sovittamiseen ja verrata sitten keskenään mallien suorituskyykyä arvioiden samalla tulosten luotettavuutta (Elith & Leathwick 2009; Qiao ym. 2015). Menetelmän valinnalla ei voida korjata huonolaatuisen aineiston tuottamaa epävarmuutta lajin ja ympäristön suhteen mallintamisessa (Jarnevich ym. 2015). Tärkeintä levinneisyysmallinnuksessa on saada mahdollisimman hyvin esiin havaintojen ja ympäristön suhde, mikä onnistuu parhaiten tarkkoilla, oikein valituilla havainnoilla ja ympäristömuuttujilla sekä tätä suhdetta hyvin mallintavalla menetelmällä (Franklin 2009; Jarnevich ym. 2015). Lopulliset kriteerit menetelmän valinnalle on käyttäjän itse määriteltävä (Franklin 2009; Guisan ym. 2017).

Käytettävien menetelmien valinta on pitkälti vaihtokauppaa kahden tärkeän mallin vaatimuksen välillä – mallin yleistettävyyys ja mallin tarkkuus (Guisan & Zimmermann 2000). Merow ynnä muut (2014) painottavat, että mallin monimutkaisuus tulee sovittaa mallinnuksen tavoitteiden mukaan. Liian yksinkertainen malli voi alisovittaa havaintoja ja ympäristöä toisiinsa, minkä tuloksena ympäristön vaikutus ei ilmene yhtä voimakkaana kuin todellisuudessa, kun taas monimutkaisten mallien yleinen riski, ylisovittaminen, tuottaa liian vahvoja, todellisuudesta poikkeavia vaikutuksia ympäristön ja lajin välille. Hyvänkin ennustuskyyvyn omaava malli voi tuottaa heikkoja ennusteita itsenäisessä aineistossa,

jos esimerkiksi ylisovittaminen on ollut voimakasta (Randin ym. 2006; Peterson ym. 2011; Hijmans 2012).

Levinneisyysmallinnuksen tulokset voivat tukea aiempia teorioita ja hypoteeseja tai luoda uusia, mutta tuloksia ei tule pitää varmistettuna todisteena lajin ja ympäristön suhteesta johtuen menetelmien korrelatiivisesta luonteesta ja havaintoihin sekä ympäristömuuttujiin kuuluvasta epävarmuudesta (Austin 2007). Toisinaan havaittu kriittinen suhtautuminen levinneisyysmallinnukseen onkin usein seurausta epätietoisuudesta kahden merkittävän ennakko-odotuksen suhteen: mitä malleilla voidaan todellisuudessa esittää ja mitä käyttäjät odottavat mallien esittävän (Araújo & Peterson 2012). Yksinkertaisimmillaan mallinnuksen tavoitteena voi olla esimerkiksi löytää tulosten perusteella lajin levinneisyyteen vaikuttavia tekijöitä (Austin 1990) tai arvioida lajin esiintymistä alueilla, joita ei ole aiemmin kartoitettu (Guisan & Zimmermann 2000). Syntyneitä malleja tai sen avulla johdettuja spatiaalisia ennusteita voidaan hyödyntää esimerkiksi tehtäessä päätöksiä lajien ja habitaattien suojelusta (Titeux 2007; Franklin 2013), ilmastonmuutoksen (Pearson & Dawson 2003; Austin 2011) tai maankäytön muutosten (Wisz 2008) vaikutuksista lajien esiintymiseen. Käytettyjen menetelmien soveltuvuus aineistoon ja tutkimuksen tavoitteisiin sekä itse aineisto määrittelevät, kuinka tuloksia voidaan hyödyntää ilmiön selittämiseen (Guisan & Thuiller 2005). Jos malli näyttää sovittavan havainnot ja ympäristön huonosti tai jos tuo suhde on ekologisesti epäuskottava ja liian monimutkainen, ei tuloksia ole mielekästä hyödyntää jatkotutkimuksissa tai ekstrapoloida uudelle alueelle tai toiseen aikaan (Elith & Leathwick 2009; Guisan ym. 2017). Mallin heikko suorituskyky voi olla seurausta aineistoon liittyvistä rajoitteista, heikosta yhteydestä lajin ja muuttujien välillä tai itse menetelmä voi olla huono kyseiselle aineistolle ja tavoitteille (Elith ym. 2006).

Levinneisyyteen vaikuttavat tekijät

Eliömaantieteilijät ovat pyrkineet määrittelemään **kolme** keskeisintä, lajin levinneisyyden muodostumisen kannalta olennaista ehtoa, jotka lajin tulee täyttää voidakseen kolonisoida uuden habitaatin ja ylläpitää sinne muodostuvaa populaatiota (Pulliam 2000; Guisan & Thuiller 2005; Soberón & Peterson 2005). Jotta uusi habitaatti on ylipäättään mahdollista kolonisoida, tulee lajin yksilöiden kyetä saavuttamaan se levittäytyen habitaattia ympäröiviltä alueilta (Barve ym. 2011). Levittäytymistä kutsutaan yleisesti myös dispersaaliksi. Habitaatin abioottisten eli elottoman luonnon muodostamien ympäristöolosuhteiden tulee olla lajin kannalta suotuisat mahdollistaen selviytymisen ja lisääntymisen alueella (Austin

2002). Abioottiset olosuhteet eivät suinkaan ole ainoita lajin esiintymiseen vaikuttavia tekijöitä habitaatissa vaan myös niin sanotun bioottisen ympäristön tulee olla lajin kannalta sopiva (Chase & Leibold 2003). Bioottisella ympäristöllä tarkoitetaan pääosin lajin ja muiden habitaatin elollisten organismien välisiä, lajin kannalta positiivisia tai negatiivisia vuovaikutussuhteita.

Lajit reagoivat abioottisen ja bioottisen ympäristön vaikutuksiin ja vaihteluun eri tavoin (Pulliam 2000). Toiset lajit ovat herkempiä erilaisten ympäristötekijöiden vaihtelulle kuin toiset, minkä seurauksena lajien levinneisyysalueiden koko voi poiketa merkittävästi (Lomolino ym. 2017). Ympäristötekijöiden vaikutus lajien levinneisyyteen vaihtelee siirryttäessä yksilöiden ja populaatioiden esiintymisen tarkastelusta kohti koko lajin levinneisyysaluetta, joka voi yleisimmillä lajeilla kattaa jopa useita maanosia (Mackey & Lindenmayer 2001; Pearson & Dawson 2003; Guisan ym. 2007; Pigot & Tobias 2013). Habitaatin abioottiset ja bioottiset tekijät ja dispersaali eivät nekään aina riitä selittämään levinneisyysalueiden muodostumista. Kahden evolutiivisen tekijän, ajan ja sattuman, roolia lajien levinneisyydessä ei voida tarkasti määritellä, mutta molemmat ovat vaikuttaneet ja vaikuttavat vastedeskin taustalla niin lajeihin kuin myös niitä ympäröivään abioottiseen ympäristöön (Lomolino ym. 2017: 59, 90, 92). Ajan ja sattuman monimutkaista roolia lajien levinneisyydessä ovat käsitelleet muun muassa Gaston (2009), Svenning ja Skov (2004) sekä Pearman ynnä muut (2008).

Dispersaali

Lajin levinneisyysalue voi kasvaa ainoastaan, jos laji kykenee kolonisoimaan uusilla alueilla sijaitsevia habitaatteja ja ylläpitämään näin muodostuvia populaatioita (Soberón 2007). Uusien alueiden kolonisointi on mahdollista vain, jos laji kykenee saavuttamaan ne (Soberón & Peterson 2005). Yksilöiden dispersaali kattaa sekä lajin liikkumisen näiden habitaattien välillä että niiden sisällä ollen näin tärkeä osa esimerkiksi populaatioiden synnyn ja kasvun kannalta. (Barve ym. 2011). Eri lajien dispersaalimenetelmät poikkeavat toisistaan, mikä vaikuttaa myös dispersaalin nopeuteen ja tehokkuuteen. Toiset lajit liikkuvat hitaasti, kun taas toisille on kehittynyt liikkumista helpottavia menetelmiä, joiden avulla pitkiäkin välimatkoja voidaan kulkea verrattain lyhyessä ajassa. Dispersaalissa käytettävä menetelmä vaikuttaa myös siihen, pystyykö laji ohittamaan erilaisia dispersaalia rajoittavia esteitä, kuten maantieteellisiä esteitä, ja miten paljon nämä esteet rajoittavat yksilöiden liikkeitä (Lomolino ym. 2017: luku 6).

Huono dispersaaliokyky rajoittaa saavutettavien habitaattien määrää ja levinneisyysalueen kokoa (Soberón 2007; Barve ym. 2011). Levinneisyysalueen ulkopuolelle voi tällöin jäädä myös hyvälaatuisia habitaatteja, joiden elinolosuhteet kyseiselle lajille olisivat suotuisammat kuin niissä habitaateissa, joita ne tässä hetkessä asuttavat (Pulliam 2000; Barve ym. 2011; Peterson ym. 2011). Nopeasti dispersoivat lajit voivat siirtyä uuteen ympäristöön lyhyessä ajassa, joten niiden potentiaalisen levinneisyysalueen voidaan sanoa olevan lyhyellä aikavälillä mitattuna hitaita lajeja suurempi (Brown ym. 1996). Erinomaisen dispersaaliokyvyn omaavat lajit, kuten lentokykyiset linnut, kykenevät teoriassa saavuttamaan pääosan potentiaalisista habitaateistaan (Soberón & Peterson 2005).

Abioottinen ympäristö

Abioottisilla ympäristötekijöillä tarkoitetaan elottoman luonnon aiheuttamia fyysisiä ja kemiallisia tekijöitä, jotka vaikuttavat eliöihin (Austin & Smith 1989). Tyypillisiä esimerkkejä ovat ilmastoon liittyvät tekijät, kuten lämpötila ja sateisuus, auringon valo ja ravinteet. Abioottiset tekijät voivat olla säätelijöitä, jotka vaikuttavat eliöiden esiintymiseen ja joiden määrään eliöt eivät vaikuta tai resursseja, joita eliöt kuluttavat (Austin 2002, 2007). Abioottiset ympäristötekijät luovat yhdessä bioottisen ympäristön kanssa paikallisia, toisistaan poikkeavia habitaatteja, joihin eliölajien on sopeuduttava menestyäkseen (Kearney 2006). Maantieteellisesti vaihtelevat ympäristöolosuhteet mahdollistavat erilaisten lajien esiintymisen maapallolla. Samoja lajeja on mahdollista tavata myös toisistaan poikkeavista abioottisista ympäristöistä (Austin & Smith 1989; Lomolino ym. 2017).

Luonnossa havaittu lajien esiintymisen vaihtelu johtuu sopeutumisesta eli adaptaatiosta tiettyihin ympäristöolosuhteisiin (Pulliam 2000). Lajeille on evoluution aikana kehittynyt erilaisia adaptaatioita eli sopeutumista helpottavia ominaisuuksia, jotka ovat mahdollistaneet lajien selviytymisen niiden historiallisissa habitaateissa (Reeve & Sherman 1993; Lomolino ym. 2017). Adaptoituminen tietynlaiseen abioottiseen ympäristöön voi vaikeuttaa selviytymistä tästä ympäristöstä poikkeavissa olosuhteissa tai kyseisen habitaatin olosuhteiden muuttuessa (Reeve & Sherman 1993; Kearney & Porter 2009). Adaptoitumista abioottisiin ympäristöolosuhteisiin tapahtuu usein myös kilpailukyvyn heikkenemisen uhalla; lajien evoluutio on jatkuvaa kompromissien tekemistä, jolloin tietynlaiseen ympäristöön erikoistuminen vaikeuttaa selviytymistä tästä poikkeavissa ympäristöissä (Kearney & Porter 2009).

Lajin fysiologia on tärkein tekijä sopeutumisen kannalta määritellessään, millaisia ympäristöolosuhteita laji kykenee sietämään (Whittaker 1967; Austin 2002). Lajin reaktiota kulloinkin vallitseviin olosuhteisiin kutsutaan fysiologiseksi vasteeksi – eri lajit reagoivat eri tavalla esimerkiksi lämpötilan tai ilmankosteuden määrään ja vaihteluun, mikä johtaa lajien esiintymisen vaihteluun erilaisissa ympäristöissä (Whittaker 1967). Sopeutuminen lämpötilan vaihteluun on erinomainen esimerkki lajien poikkeavista fysiologisista vasteista ja sen aiheuttamasta jakautumisesta maapallolla (Pörtner ym. 2006). Ympäristöolosuhteiden ollessa optimaaliset yksilön kasvu ja lisääntyminen ovat teoriassa tehokkaimmillaan (Austin & Smith 1989). Siirryttäessä tästä optimaalisesta ympäristön tilasta pois päin kohti huonompia olosuhteita lajin suorituskyky heikkenee, kunnes olosuhteet ovat sellaiset, joissa pidempiaikainen oleskelu käy mahdottomaksi (Pörtner ym. 2006). Tämä siirtymä optimaalisesta heikkoon suorituskykyyn voi tapahtua nopeasti tai hitaasti riippuen lajin sietokyvystä ympäristötekijän vaihteluun (Austin 2002).

Koska abioottinen ympäristö muodostuu ympäristötekijöiden muodostamasta kokonaisuudesta, eliöt altistuvat samanaikaisesti kaikkien näiden tekijöiden yhteisvaikutukselle (Kearney & Porter 2009). Tällöin yksittäisten ympäristömuuttujien vaikutusten arvioiminen lajin esiintymiseen voi muodostua mahdottomaksi (Austin 2007). Samassa ympäristössä voi olla sekä lajin esiintymistä rajoittavia että sitä helpottavia ympäristötekijöitä, ja eri tekijät voivat vahvistaa tai heikentää toistensa vaikutusta (Hirzel & Le Lay 2008). Tietty ympäristögradientti ei välttämättä yksinään estä lajin selviytymistä ympäristössä eivätkä yksittäisten tekijöiden vaikutukset ole aina niin voimakkaita kuin ajatellaan (Guisan & Thuiller 2005). Useat lauhkean ilmastovyöhykkeen lajit esimerkiksi pystyvät sietämään verrattain lyhytaikaista kylmyyttä ilman suurta vaikutusta yksilöiden selviytymiseen, mutta samojen seutujen lyhyt kasvukausi rajoittaa kasvien perustuotantoa ja sitä myöten resursien saatavuutta, mikä rajoittaa koko eliöyhteisön lajien lukumäärää (Root 1988).

Abioottiset ympäristötekijät, jotka ovat lajin vähimmäisvaatimusten rajoissa mahdollistaen lajin selviytymisen, muodostavat lajin ekologisen lokeron (*ecological, environmental niche*) rajat (Austin 2002; Chase & Leibold 2003; Peterson ym. 2011). Lajin ekolokero tarkoittaa tätä abioottisen ympäristön ja lajin ominaisuuksien määrittelemää lajin paikkaa ympäristössä, mikä ilmenee lopulta lajin esiintymisessä tilassa ja ajassa. Hutchinson (1957) määritteli aiempien ekologien ajatusten pohjalta lajeille ympäristöolosuhteiden rajaaman peruslokeron (*fundamental niche*), joka kattaa lajin esiintymisen mahdollistavan abioottisen ympäristön. Muiden lajien ja dispersaalin vaikutuksesta peruslokero rajautuu

niin kutsutuksi toteutuneeksi lokeroksi (*realized niche*), joka sitten näyttäytyy havaitsemasamme lajin levinneisyydessä ja runsaudessa (Soberón 2007). Hutchinsonin ajatuksia lajin ekolokerosta on myöhemmin muokattu lukuisten tutkijoiden toimesta, mutta perusajatus on pysynyt samana (Chase & Leibold 2003; Kearney 2006; Soberón 2007). Lajien levinneisyyden muodostumista ajatellen ekolokero on hyödyllinen käsite, sillä se sisältää lajin vaatimat ympäristöolosuhteet ja resurssit sekä määrittelee ympäristöt, joissa laji kykenee esiintymään. Samassa ympäristössä elävät lajit altistuvat samojen ympäristötekijöiden vaikutuksille (Kearney 2006). Lajin ominaisuudet määrittelevät, kuinka laji reagoi näihin tekijöihin ja sen, millaiseksi lajin ekolokero muodostuu.

Bioottinen ympäristö

Habitaatin saavuttaminen, ja sopeutuminen abioottiseen ympäristöön ovat tärkeitä tekijöitä uusien habitaattien kolonisoinnin kannalta, mutta olennainen merkitys on myös sillä, millainen asema lajilla on osana kyseisen ekosysteemin eliöyhteisöä (Austin 2002; Chase & Leibold 2003). Lajien väliset eli bioottiset vuorovaikutussuhteet ovat tärkeitä lajien selviytymisen, kasvun ja lisääntymisen kannalta, sillä ne altistuvat väistämättä toistensa toiminnan vaikutuksille sekä suoraan että epäsuorasti (Austin 2002; Pearson & Dawson 2003). Muut lajit voivat vaikeuttaa tai helpottaa lajin esiintymistä ja populaatioiden ylläpitämistä erilaisissa habitaateissa (Araújo & Luoto 2007; Pigot & Tobias 2013). Bioottinen ympäristö saattaa myös vaikuttaa lajin sopeutumiseen eri tavoin eri osissa ympäristögradienttia (Wisz ym. 2013).

Negatiivisista bioottisista tekijöistä tyypillisin on kilpailu samoista, usein rajallisista resursseista, kuten ravinnosta ja elintilasta (Gause 1934; Leibold 1995; Meier ym. 2010). Resurssien saatavuus vaikuttaa yksilöiden selviytymiseen, kasvuun ja lisääntymiseen ja sitä kautta myös populaatioiden kokoon ja määrään sekä levinneisyyteen (Wisz ym. 2013). Lajinsisäinen ja lajienvälinen kilpailu heikentää lajien mahdollisuutta hyödyntää ympäristöolosuhteiltaan sopivia habitaatteja (Connell 1961) kaventaen näin niiden potentiaalista esiintymisaluetta (Soberón 2007; Pigot & Tobias 2013). Kilpailijat voivat kaventaa sen toteutunutta lokeroa tai jopa syrjäyttää lajin sille optimaalisesta abioottisesta ympäristöstään (Austin & Smith 1989; Austin 2002).

Kilpailun rinnalla negatiivisia vaikutuksia lajin kannalta on myös esimerkiksi vaara tulla hyödynnetyksi ravintona muiden lajien toimesta (Leibold 1995) ja taudinaiheuttajilla ja loisilla (Lira-Noriega & Peterson 2014). Predaatiossa eli saalistuksessa toinen

eläinlaji, peto, hyödyntää toista eläinlajia ravintonaan (Leibold 1995); peto-saalis -suhde voi vaikuttaa sekä saaliseläimen että pedon paikalliseen runsauteen ja esiintymiseen jopa laajemmassa mittakaavassa (Araújo & Rosenfeld 2014). Lajiin kohdistuva voimakas saalistuspaine voi pienentää sen levinneisyysaluetta esimerkiksi sellaisessa tapauksessa, jossa lajin populaatiokoot pienenevät predaation johdosta johtaen paikallisiin ekstinktioihin eli lajin häviämisiin tietyiltä alueilta (Holt & Barfield 2009). Pedon ja saalislajien välinen vuorovaikutus on tärkeää myös pedon näkökulmasta: ravintona käytettävien lajien kannat ohjaavat usein myös petojen populaatioita ja esiintymistä ajassa ja paikassa (Leibold 1995; Holt & Barfield 2009). Tätä vuorovaikutusta voi laajentaa koskemaan myös kasvinsyöjien eli herbivorien ja kasvien välistä suhdetta, joka voi vaikuttaa molempien esiintymiseen (Wisz ym. 2013).

On olemassa myös sellaisia bioottisia tekijöitä, jotka voivat helpottaa lajin sopeutumista abioottisen tekijöiden ja lajien muodostamaan ympäristöön (Boucher ym. 1982; Bruno ym. 2003; Erwin 2008). Kahden lajin rinnakkaiselo, jossa molemmat hyötyvät toisensa läsnäolosta kutsutaan mutualismiksi (Boucher ym. 1982). Ilmeinen mutualistinen vuorovaikutus on kasvien ja niiden pölyttäjien välinen, molempia hyödyttävä suhde. Läheskään kaikki mutualistiset suhteet eivät ole näin ehdottomia lajien esiintymisen kannalta; on esimerkiksi mahdollista, että useampi laji kykenee täyttämään saman roolin, kuten on useimpien hyönteispölytteisten kasvien kohdalla (Waser ym. 1996). Habitaateissa voi esiintyä lajeja, jotka edesauttavat muiden lajien esiintymistä kyseisessä ympäristössä (Bruno ym. 2003) esimerkiksi muokkaamalla abioottisen ympäristön olosuhteita muille lajeille hyödyllisiksi, parantamalla resurssien saatavuutta ja luomalla uusia mikrohabitaatteja (Erwin 2008; Linder ym. 2012). Lajille hyödylliset bioottiset vuorovaikutussuhteet voivat lisätä suotuisten habitaattien määrää ja laatua (Heikkinen ym. 2007) ja laajentaa lajin esiintymisaluetta (Araújo & Rosenfeld 2014).

Predaation, kilpailun tai mutualismin kaltaisten bioottisten vuorovaikutussuhteiden aiheuttamia todellisia muutoksia lajin populaatioiden esiintymisessä on kuitenkin vaikea havaita luonnossa ilman intensiivistä tutkimusta, kuten keinotekoisia tutkimusasetelmiä, joissa hyödynnetään lajien siirtoa luonnollisten levinneisyysalueiden ulkopuolelle (Elith & Leathwick 2009; Kissling ym. 2012). Bioottisten vuorovaikutussuhteiden ja abioottisen tekijöiden luoma ympäristö on monimutkainen kokonaisuus, jossa yksittäisten tekijöiden rooli lajin selviytymisen ja esiintymisen kannalta voi olla vaikea tunnistaa (Guisan & Thuiller 2005; Araújo & Guisan 2006; Soberón 2007). Samojen lajien väliset

vuorovaikutukset voivat poiketa erilaisissa ympäristöissä, ja ympäristön muutokset voivat muuttaa vallitsevien bioottisten suhteiden muotoa (Wisz ym. 2013). On myös hyvinkin mahdollista, että lajia rajoittavat eri puolilla sen levinneisyysaluetta myös erilaiset tekijät (Brown ym. 1996): levinneisyysalueen raja voi yhtäällä muodostua ympäristötekijöiden seurauksena ja toisaalla muiden lajien rajoittavan tai positiivisen vaikutuksen vuoksi (Austin 1990). Lajien väliset suhteet muodostavat erilaisia verkostoja, jotka kytkevät saman habitaatin lajeja toisiinsa välillisesti myös ilman suoranaista kontaktia (Kissling ym. 2012; Anderson 2017). Tällöin koko verkoston toiminta voi vaikuttaa yksittäisten lajien esiintymiseen – esimerkkinä ekosysteemien ravintoverkot (Ings ym. 2009).

Päiväpetolinnut Suomessa – uhanalaistuvien lajien määrä on kasvussa

Suomessa pesii vakituisesti lukuisia päiväpetolinnuiksi luokiteltavia lajeja (Gensbøl ym. 1995). Suomessa tavattavia lajeja esiintyy niin muualla Euraasiassa kuin myös Pohjois-Amerikassa eikä yhdenkään lajin levinneisyysalue ole rajoittunut esimerkiksi pelkkään pohjoiseen Eurooppaan (The IUCN Red List... 2020). Useiden lajien levinneisyysalueen pohjoisraja taas rajoittuu Suomen leveysasteille (The IUCN Red List... 2020), mikä tekee Suomesta kiinnostavan alueen päiväpetolintujen levinneisyyden tarkasteluun. Pääosa Pohjois-Euroopan päiväpetolinnuista on muuttolintuja, jotka talvehtivat esimerkiksi Etelä-Euroopassa tai Afrikassa (Gensbøl ym. 1995). Vain muutamat lajit viettävät koko vuoden Suomessa.

Päiväpetolintuja tavataan monipuolisissa ympäristöissä, esimerkiksi metsissä, soilla, viljelyalueiden ja vesistöjen läheisyydessä ja avoimilla tunturikankailla (Gensbøl ym. 1995; Lehikoinen ym. 2019). Habitaatin valintaa ohjaavat lajin pesimäympäristövaatimukset ja ravinnonhankintamenetelmät (Gensbøl ym. 1995). Päiväpetolinnut käyttävät ravintonaan monipuolisesti erilaisia jysijöitä, lintuja ja muita pieneläimiä, joiden habitaatit vaihtelevat. Huolimatta saaliseläimiksi kelpaavien lajien määrästä useat lajit ovat ruokavaliossaan erikoistuneet tiettyyn lajiin tai eliöryhmään ja kyseisen ravinnon saatavuus ohjaa petolintujen habitaatin valintaa ja pesinnän onnistumista eri vuosina. Useimmat lajit kuitenkin kykenevät hyödyntämään vaihtoehtoisia saaliseläimiä vuosina, joina ensisijaisen ravinnon määrä on vähäinen.

Useat aiemmin yleiset päiväpetolintulajit ovat uhanalaistuneet Suomessa lyhyessä ajassa (Lehikoinen ym. 2019). Myös näiden lajien maailmanlaajuiset kannat ovat laskussa

(The IUCN Red List... 2020). Suurimmat syyt ja tulevaisuuden uhat päiväpetolinnuille, kuin myös muullekin maailman lajistolle, ovat ilmastomuutos ja ihmisen toiminnan seurauksena tapahtuvat maankäytön muutokset, kuten metsien ja soiden häviäminen sekä maatalouden muuttuminen (Lehikoinen ym. 2019). Lajien luonnollisten habitaattien katoamisen ohella päiväpetolintuja ovat uhanneet myös ihmisten harjoittama vaino, häirintä ja munien kerääminen (Gensbøl ym. 1995; Lehikoinen ym. 2019). Muuttolintujen kohdalla vaarana on myös talvehtimisalueiden ympäristössä tapahtuvat muutokset (Lehikoinen ym. 2019), itse muuttomatka keväällä ja syksyllä (Laaksonen & Lehikoinen 2013; Klaassen ym. 2014) sekä ilmastomuutoksen aiheuttamat muutokset muuton ajankohdassa (Saino ym. 2011). Ilmastomuutos vaikuttaa sekä suoraan päiväpetolintujen habitaatin abioottisiin olosuhteisiin että bioottisiin vuorovaikutussuhteisiin (Lehikoinen ym. 2019). Pesimisen ajankohdan väärä ajoitus, vaihteleva sää ja esimerkiksi sateiden lisääntyminen voivat aiheuttaa pesinnän epäonnistumista keväällä ja kesällä (Lehikoinen ym. 2009). Vuodenaikaisten lämpötilojen muutokset ja lumipeitteen väheneminen vaikuttavat petolinnuille tärkeiden jyrsijöiden, kuten myyrien, runsauteen ja runsauden vaihteluun vuosien välillä muodostaen uhan herkkien petolintupopulaatioiden selviytymiselle (Gilg ym. 2009; Schmidt ym. 2012; Korpela ym. 2013; Terraube ym. 2015). Ilmastomuutos mahdollistaa eteläisten lajien siirtymisen kohti pohjoista, mikä voi johtaa esimerkiksi lajienvälisen kilpailun lisääntymiseen ja Suomen lajien taantumiseen, erityisesti pohjoisen ankariin olosuhteisiin sopeutuneissa lajeissa (Chen ym. 2011). Muuttuva ympäristö, ja sen aiheuttamat muutokset ravintoverkostojen tasapainossa sekä lisääntynyt kilpailu ylittävät todennäköisesti lämpötilojen nousun ja kasvukauden pidentymisen hyödyt pohjoisille leveyksille sopeutuneiden lajien näkökulmasta (Terraube ym. 2015).

Uhanalaistuminen ei rajoitu tiettyihin habitaatteihin – esimerkkilajeina piekana ja mehiläishaukka

Viime vuosien aikana uhanalaistuneita päiväpetolintuja tavataan jokaisessa habitaatissa soista ja pelloista metsiin ja tuntureille, pohjoisesta etelään, ravinnoltaan monipuolisista generalistilajeista pitkälle erikoistuneisiin specialisteihin (Laaksonen & Lehikoinen 2013; Björklund ym. 2015; Lehikoinen ym. 2019). Yksi vaarantuneista lajeista on piekana (Terraube ym. 2015). Piekanan levinneisyysalueen pääpaino on Lapissa, mutta laji pesii toisinaan myös esimerkiksi Länsi-Suomessa Pohjanmaan rannikolla hyvinä myyrävuosina (Gensbøl ym. 1995; Suomen Lajitietokeskus 2020). Piekanan ravintona käyttämien

jiysijöiden, kuten myyrien, määrä voi vaihdella vuosien välillä runsaasti eri puolilla maata ja saaliseläinten ohjaavat piekanan esiintymistä ja pesinnän onnistumista (Gensbøl ym. 1995; Reif ym. 2001). Huonoina myyrävuosina piekana saalistaa esimerkiksi päästäisiä ja kana-, vesi- sekä pikkulintuja (Reif ym. 2001). Laji suosii pesimäpaikakseen tuntureiden rinteiden kallionseinämiä pohjoisessa ympäristössään – etelämpänä yksilöt pesivät harva- puustoisissa mäntymetsissä (Gensbøl ym. 1995: 168–173). Saalistusmaastona ne puolestaan suosivat avoimia luonnonympäristöjä, kuten tunturikankaita ja muita aukeita alueita, joissa jirsijöitä on helpompi saalistaa. Piekanat talvehtivat Suomen ulkopuolella ja niiden paluumuutto tapahtuu huhti-toukokuussa. Kevät on kriittistä aikaa päiväpetolintujen muuton ja pesinnän kannalta – kevään nopea lämpeneminen verrattuna muihin vuodenaikoihin voi helpottaa kilpailevien lajien sopeutumista piekanan pesimäseuduille (Lehikoinen ym. 2009)

Mehiläishaukan esiintyminen painottuu vahvasti Etelä- ja Keski-Suomeen eikä lajia tavata säännöllisesti esimerkiksi Lapissa (Suomen Lajitietokeskus 2020). Mehiläishaukan suosimia habitaatteja ovat esimerkiksi runsaan kasvillisuuden omaavat, vanhat lehti- ja sekametsät, pohjoisempina myös rehevät havumetsät (Amcoff ym. 1994; Gensbøl ym. 1995; Selås 1997). Lajin ravintokäyttäytyminen on päiväpetolinuille poikkeuksellinen sillä pääosa niiden ravinnosta koostuu mesi- ja petopistiäisistä, kuten mehiläisistä ja ampiaisista sekä muista hyönteisistä (Itämies & Mikkola 1972; Gensbøl ym. 1995). Hyönteisravinnon suosiminen on luultavasti suuri syy sille, että lajin levinneisyys painottuu Suomessa eteläiseen Suomeen sillä hyönteisten määrä kasvaa keväisin ja kesäisin vasta, kun sää lämpenee riittävästi. Mehiläishaukat eivät myöskään viihdy kauaa viileillä pesimäseuduillaan Suomessa: kevätmuutto tapahtuu vasta touko-kesäkuun aikana, kun taas syysmuutto talvehtimisalueille alkaa jo loppukesästä, hyvissä ajoin ennen säiden viilentymistä (Gensbøl ym. 1995: 56–60). Vaikka Suomen lämpeneminen voisikin lisätä mehiläishaukalle sopivan saaliin ja saalislajien määrää alkukesästä, lajin kannat ovat pienentyneet todennäköisesti vanhojen metsien ja perinnebiotooppien katoamisen myötä (Björklund ym. 2015; Lehikoinen ym. 2019). Suomalaisista mehiläishaukoista on löydetty myös neonikotinoidien, viljelyssä käytettävien hyönteismyrkköjen pitoisuuksia (Byholm ym. 2018).

Käsitteellisestä mallista kohti levinneisyysmallinnusta

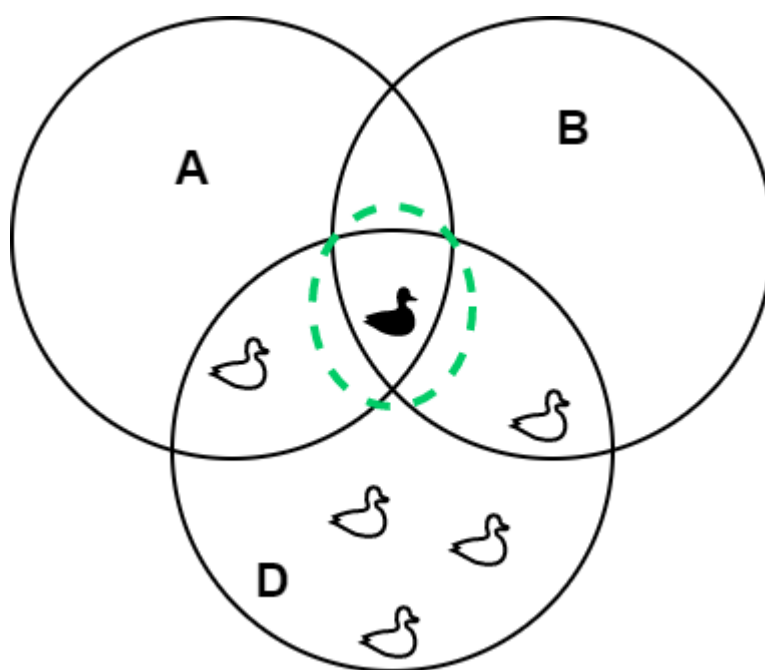
Tavoitteena ennustaa lajin potentiaalista tai toteutunutta levinneisyysaluetta

Levinneisyysmallinnuksessa on perinteisesti hyödynnetty ekologien määrittelemää ekologikon käsitettä, ja siihen liittyvää teoreettista näkökulmaa lajin ympäristövaatimuksista ja ympäristön aiheuttamista rajoitteista levinneisyysalueiden muodostumisesta (Pulliam 2000; Austin 2002). Tutkimusalan ja menetelmien kehittyessä kritiikki ekologikoteoriaan tukeutumisesta lajien levinneisyyden mallinnuksessa on vahvistunut (Guisan & Thuiller 2005; Hirzel & Le Lay 2008). Esimerkiksi, pitkään vallalla ollut yksinkertaistus, jonka mukaan levinneisyysmallinnuksen tuloksena saadaan määriteltyä Hutchinsonin toteutunut lokero, on sitemmin kyseenalaistettu (Guisan & Thuiller 2005; Franklin 2009). Kritiikki ja keskustelu ovat johtaneet lukuisiin ekologikon ja levinneisyyden suhdetta koskeviin teorialähtöisiin julkaisuihin (Chase & Leibold 2003; Kearney 2006; Soberón 2007; Peterson ym. 2011), jotka käsittelevät syvällisemmin ekologikon ja levinneisyyden suhdetta pyrkien havainnollistamaan ekologikon ilmenemistä ympäristössä ja maantieteellisessä tilassa. He osoittavat, että näitä toisilleen lähekkäisiä käsitteitä on käytetty huolimattomasti päällekkäin ja että rajanveto käsitteiden välillä on välttämätön. Termi ekologikero sisältää ajatuksen populaation kasvusta, minkä vuoksi asutetun habitaatin olosuhteiden tulee olla lajin lisääntymiselle suotuisat, jotta siitä voidaan puhua osana lajin ekologista ekologikeroa (Chase & Leibold 2003). Lajin levinneisyys ei kuitenkaan aina sisällä kaikkia niitä habitaatteja, jotka ovat abioottiselta ja bioottiselta ympäristöltään sopivia lajin luonnollisten populaatioiden ylläpidolle eli lajin ekologista lokeroa (Pulliam 2000; Soberón 2007; Peterson ym. 2011). Luonnossa havaitsemamme lajin levinneisyyden voidaan katsoa heijastavan spatiaalisesti sen ekologikeroa maantieteellisessä tilassa (Lomolino ym. 2017: 78–79). Abioottisen ja bioottisen ympäristön rajoittavat tekijät määrittelevät, millaisella maantieteellisellä alueella lajia voidaan tavata (Austin 2002; Peterson ym. 2011).

Pulliam (2000) hyödynsi ekologikoteorian rinnalla lähde-nielu-mallia (Pulliam 1988) ja metapopulaatioteoriaa (Hanski 1999) selittäessään lajien maantieteellisen levinneisyyden muodostumista. Hän esitti, että ekologikeroista riippumattomat tekijät, kuten dispersaalin rajoitteet ja populaatiodynamiikan aiheuttama vaihtelu lajirunsaudessa voivat estää esiintymisen ympäristössä, joka abioottiselta ja bioottiselta ympäristöltään olisi muutoin sopiva. Samoin on mahdollista, että lajin yksilöitä tavataan ympäristössä, jonka

olosuhteet – abioottinen tai bioottinen ympäristö tai molemmat – eivät ole suotuisat populaatiokoon ylläpidolle. Tällaisen nielupopulaation säilyminen habitaatissa on täysin riippuvainen ympäröivistä lähdepopulaatioista ja niiden koosta.

Pulliamin (2000) näkemys on ollut usein pohjana teoreettisissa malleissa, jotka pyrkivät kuvaamaan levinneisyysalueiden muodostumista maantieteellisessä tilassa (kuva 2, Soberón & Peterson 2005; Peterson ym. 2011). Malleissa määritellään lajin toteutuneeksi levinneisyydeksi se alue, jolla laji esiintyy. Maantieteellinen levinneisyysalue voi sisältää sekä ympäristöltään sopivia että epäsopivia habitaatteja (Pulliam 2000; Soberón 2007). Suotuisan ympäristön sisältävä alue kykenee ylläpitämään metapopulaatiokokoja vaikuttaen niiden väliseen dynamiikkaan ja nielupopulaatioiden määrään. Alueiden asuttamista rajoittavat lajin dispersaaliokyky ja levinneisyysesteet (Soberón 2007). Ympäristöltään sopivia habitaatteja, joiden voidaan sanoa kuuluvan lajin potentiaaliseen levinneisyysalueeseen (Gaston 2009; Soberón & Peterson 2005), voi jäädä kolonisoimatta, jos laji ei kykene ylittämään näitä rajoitteita (Peterson ym. 2011).



Kuva 2. Lajin levinneisyysalue muodostuu suotuisan abioottisen (A) ja bioottisen (B) ympäristön ja dispersaalilla saavutettavan alueen (D) yhdistämään maantieteelliseen tilaan (Soberón & Peterson 2005, muokkauksin). Vihreällä katkoviivalla korostettu alue kuvaa tilaa tai habitaattia, joka on lajin saavutettavissa ja abioottiselta ja bioottiselta ympäristöltään riittävä ylläpitämään ja kasvattamaan populaatiokokoa.

Habitaatin asuttaminen ei siis suinkaan tarkoita sitä, että se olisi lajin kannalta sopiva (Van Horne 1983; Pulliam 2000; Soberón 2007). Levinneisyysmallinnuksessa on usein lähtökohtana ajatus, että laji on tasapainossa ympäristönsä kanssa, minkä seurauksena levinneisyysalueessa ja populaatioissa tapahtuisi ainoastaan pieniä muutoksia lyhyellä tarkasteluvälillä (Austin 2002; Guisan & Thuiller 2005). Näin lajiin oletettavasti vaikuttavien ympäristötekijöiden ja yksilöhavaintojen yhdistämisellä voisi teoriassa mallintaa sekä lajin levinneisyyttä että sen ekolokeroa. Useissa tapauksissa tasapainoa lajin ja sen asuttaman ympäristön välillä ei kuitenkaan voi odottaa (Guisan ym. 2014). Pääosa levinneisyyden mallintamiseen käytetyistä malleista perustuu kuitenkin pelkkiin yksilöhavaintoihin ilman tietoa populaation rakenteesta ja sen vaihtelusta (Franklin 2009; Kissling ym. 2012). Olennainen tieto lajin populaatiodynamiikasta alueella jää usein saamatta, koska havainnot riittävän pitkän aikavälin muutoksista populaatiokoon vaihtelussa vaativat suuremman työpanoksen ja ovat toteutettavissa vain verrattain pienillä tutkimusalueilla (Elith & Leathwick 2009). Tutkimusalueen ulkopuolelle jäävien populaatioiden vaihtelu voi olla erilaista esimerkiksi erilaisen abioottisen ja bioottisen ympäristön vuoksi (Pulliam 2000; Holt ym. 2005). Populaatioon liittyvien muutosten havaitseminen olisi elintärkeää, jos havaintojen perusteella halutaan tehdä päätelmiä lajin ekolokerosta (Soberón 2007).

Useat tutkijat ovatkin pyrkineet tietoisesti tekemään eroa ekolokeron ja levinneisyyden mallintamisen välille (Kearney 2006; Hirzel & Le Lay 2008; Jiménez-Valverde ym. 2008). Huolimatta keskustelusta koskien ekolokeron ja levinneisyyden erottamista toisistaan, yksimielisyyttä on löytynyt sen suhteen, että lajien levinneisyyden taustalla vaikuttavat niiden ympäristövaatimukset ja monimutkaiset vuorovaikutussuhteet ympäristön ja muiden lajien kanssa – siis samat tekijät kuin alkuperäisissä ekolokeroteorioissa (Pulliam 2000; Austin 2002; Guisan & Thuiller 2005; Soberón 2007; Peterson ym. 2011). Yksilöiden esiintymisestä saadut havainnot yhdistettynä lajiin vaikuttavia ympäristötekijöitä kuvaaviin muuttujiin kertovat lajin toteutuneesta levinneisyydestä tai potentiaalisesta levinneisyydestä (Soberón 2007; Jiménez-Valverde ym. 2008) tietyllä alueella, tietyssä ajankohdassa (Guisan ym. 2017). Tutkimusasetelma ja valitut menetelmät ohjaavat tuloksen luonnetta ja muodostuvaa spatiaalista ennustetta, karttaa. Lajin havaintojen ja poissaolojen pohjalta johdetun mallin tuottama ennuste on esimerkiksi jatkuva indeksi habitaatin soveltuvuudesta lajille tai todennäköisyydestä tavata lajin yksilöitä kyseisessä havaintopisteessä (Guisan ym. 2017: 44).

Levinneisyysmallinnus pohjautuu lajin yksilöiden esiintymiseen

Esiintymishavaintojen ja käytettyjen muuttujien valinta tulee tehdä tutkimuksen tavoitteen ja tutkimusalueen perusteella (Franklin 2009; Elith & Leathwick 2009). Kerätty havaintoaineisto ja sitä selittämään käytettävät muuttujat, millaisia malleja voidaan käyttää. Havaintojen keräämiseen ja muuttujiin liittyvät ongelmat myös heijastuvat koko mallinnusprosessiin (Elith & Leathwick 2009). Levinneisyysmallinnuksessa käytettävään aineistoon ja tilastollisiin menetelmiin liittyykin lukuisia ongelmakohtia ja riskejä, jotka vaikuttavat prosessin eri vaiheissa ja uusia, vaihtoehtoisia menetelmiä mallintamiseen, mallien tulosten arvioimiseen ja ongelmien ratkaisuun etsitään jatkuvasti (mm. Guisan & Zimmermann 2000 Hirzel & Le Lay 2008; Jiménez-Valverde ym. 2008; Franklin 2009; Guisan ym. 2017).

Jotta lajin yksilöistä tehtyjä havaintoja voidaan hyödyntää lajin levinneisyyden mallinukseen, on niiden oltava georeferoituja eli liitetty esimerkiksi maantieteellistä aluetta kuvaavaan karttaan. Levinneisyysmallinnuksessa tarkasteltava alue jaetaan tyypillisesti hilauihin (*grid*), joilla on sama resoluutio eli solukoko, esimerkiksi 1 km x 1 km. Hilat muodostavat näin koko alueen kattavan ruudukon (rasteri) ja yksittäisiin hiloihin kirjataan sen sijainti x–y -koordinaatistossa sekä tieto, löytyykö sen sisältä kyseisen lajin yksilöitä vai ei (Guisan & Zimmermann 2000; Franklin 2009). Havaintojen esittäminen hilamuodossa mahdollistaa levinneisyyden mielekkään mallintamisen ja peittää yksilöiden paikannuksessa tapahtuneita virheitä (Franklin 2009: 27). Samalla peitetään valitettavasti myös vaihtelua lajin yksilöiden esiintymisessä eri habitaateissa (McPherson ym. 2006).

Yksinkertaisimmillaan tämä levinneisyysmallinnuksessa käytettävä lajitieto on havaintoja lajin yksilöistä valitun alueen sisällä (*presence-data*, *presence-only -data*). Tällaiset havainnot kerätään usein valmiista lajitietokannoista tai museonäytteistä eivätkä ne ole yleensä peräisin suunnitelluista kartoituksista (Guisan & Thuiller 2005; Elith & Leathwick 2009). Vaihtoehtoisesti havaintoaineistossa voi olla esiintymisten lisäksi mukana tiedot myös havaintopisteistä, joissa lajia ei ole tavattu (*presence-absence-data*); tällainen aineisto on pääosin peräisin suunnitelmallisesti toteutetuista kartoituksista, joissa koko tutkimusalue on tarkastettu havaintojen varalta tietynlaisen järjestelmällisen suunnitelman pohjalta (Guisan & Thuiller 2005; Elith & Leathwick 2009). Näin voidaan tehdä perustellumpi oletus siitä, missä lajia tavataan ja missä ei. On kuitenkin yleistä, että yksilöitä jää havaitsematta huolellisestakin kartoituksesta huolimatta: havaintojen luotettavuuteen liittyvä epävarmuus onkin suuri ongelma, kun mallinnuksen tavoitteena on löytää

yhteys havaintojen ja ympäristön välillä (Hirzel & Le Lay 2008). Yksilöiden havainnointiin liittyvää epävarmuutta voidaan yrittää korjata jälkikäteen esimerkiksi suorittamalla malliin valikoituvien havaintojen uudelleenotanta erilaisia vaihtoehtoja hyödyntäen (Fourcade 2014). Toisinaan tutkimusalueelta on havaintojen lisäksi tallennettu myös tieto tavattujen yksilöiden määrästä, jota voidaan hyödyntää eri tarkoituksiin (Guisan & Thuiller 2005). Tällaisen havaintoaineiston käyttäminen poikkeaa esiintymishavaintojen ja poissaolojen käytöstä esimerkiksi mallinnuksessa käytettäviltä menetelmiltään ja mallin tuloksilta (Renner ym. 2015).

Optimaalinen tilanne olisi, että mallinnusta varten olisi käytössä huolellisesti kerättyä, *presence-absence* -muotoista aineistoa, koska pääosa hyödynnettävistä tilastollisista menetelmistä, kuten regressiomenetelmät, ei kykene luomaan mallia pelkkien esiintymisten pohjalta (Guisan & Zimmermann 2000; Elith ym. 2006). Lajin esiintymisestä ja poissaoloista huolella kerätty otos, joka kattaa koko mallinnuksessa käytettävän alueen, tarjoaa mahdollisuuden arvioida lajin yleisyyttä eli kuinka suuren osan se asuttaa tarkasteltavasta alueesta (Elith & Leathwick 2009). Lajin yleisyys otoksessa vaikuttaa mallien tarkkuuteen. Mallien luominen käyttäen pelkkiä yksilöhavaintoja voi poistaa aineistoon liittyviä ongelmia, kuten tilanteissa, joissa lajin havaintoja ja poissaoloja kerätessä aineistonkeruu on puutteellista, tai kun populaatiodynamiikan vaihtelu on poistanut lajin hetkellisesti suotuisasta ympäristöstä (Elith & Leathwick 2009). Pelkkiä yksilöhavaintoja sisältävät aineistot kuitenkin sisältävät tyypillisemmin kartoituksen suunnitelmallisuuden puutteesta johtuvia ongelmia, kuten havaintojen painottuminen helposti saavutettaviin kohteisiin, esimerkiksi teiden ja asutuksen lähelle (Barry & Elith 2006), ja niiden pohjalta luodut mallit suoriutuvat menetelmien vertailussa pääosin huonommin kuin tunnettuja poissaoloja hyödyntävät mallit (Elith ym. 2006; Franklin 2009; Guisan ym. 2017). Jos poissaolot ovat luotettavia, niitä tulee käyttää mallinnuksessa pelkän havaintotiedon sijaan (Guisan ym. 2017).

Koska lajin poissaolon varmistaminen tietystä ympäristöstä on vaativaa ja mahdollista vain rajallisen kokoisilta tutkimusalueilta, mallinnuksessa joudutaan hyödyntämään usein pelkkiä havaintoja lajin yksilöistä (Pearce & Boyce 2006). Jos luotettavaa tietoa lajin puuttumisesta alueelta ei ole, joudutaan käyttämään menetelmiä, jotka kykenevät hyödyntämään pelkkiä esiintymishavaintoja (Guisan ym. 2017: 131) tai luomaan niin sanottua tausta-aineistoa (*background data*) korvaamaan tieto poissaoloista mallinnusta varten (Phillips ym. 2009). Tausta-aineistoksi valitaan tarpeeksi lähellä lajihavaintoja sijaitseva alue, joka kuvaa lajin saatavilla olevaa ympäristöä: näin on mahdollista saada otos

ympäristömuuttujista esimerkiksi alueelta, jolla havaintoja lajista ei ole tehty ja verrata sitä ympäristöön, josta yksilöitä on löydetty (Franklin 2009: 71). Tästä aineistosta käytetään toisinaan myös termiä *pseudo-absence*, mikä viittaa enemmän oletukseen, että valittu tausta-aineisto on sellaisesta ympäristöstä, jonka uskotaan olevan lajille epäsopeva (Phillips ym. 2009).

Tausta-aineistoksi hyödynnettävien alueiden, ja niiden lukumäärän määrittäminen mallinnusta varten on haastavaa (Barbet-Massin ym. 2012). Alueet, joilta tausta-aineisto kerätään, voidaan valita ekologisin perustein esimerkiksi siten, että ympäristöolosuhteiden uskotaan olevan alueella epäsopevat (Chefaoui & Lobo 2008) tai että päinvastoin valitaan alue, jolla lajia odotettaisiin tavattavan havaintojen puutteesta huolimatta (Phillips ym. 2009). Tausta-aineiston hyödyn ja todellisen merkityksen varmistamiseksi olisi tärkeää, että aineisto on peräisin alueelta, jonka laji kykenee saavuttamaan (Barve ym. 2011). Ilman riittävää tietoa ja varmuutta siitä, miltä alueilta tausta-aineiston otanta tulisi suorittaa, voidaan käyttää myös yksinkertaista satunnaisotantaa (Phillips ym. 2009). Satunnaisotannan voi suorittaa myös tietynlaisella painotuksella tausta-aineiston soveltuvuuden parantamiseksi (Gallien ym. 2012). Sekä ekologisin että tilastotieteellisin perustein valittu aineisto voi vahvistaa yksilöhavaintojen keruussa aiheutunutta vinoumaa, kuten puuttuvia havaintoja ympäristössä, jossa laji todellisuudessa esiintyy (Barbet-Massin ym. 2012). Guisan ynnä muut (2017: 133–134) suosittelevat käyttämään satunnaisotantaa, sillä se ei tee oletuksia aineiston perusteella ja toimii tilanteissa, joissa perusteita lajin tausta-aineiston rajaamiselle tietyille alueille ei ole (Barbet-Massin ym. 2012). On tärkeää rajata huolellisesti alue, jolta tausta-aineisto kerätään ja määritellä käytetyn menetelmän perusteella tausta-aineiston määrä suhteessa havaintojen määrään (Phillips ym. 2009): liian suuri alue ja liiallinen määrä lajia ympäröivää ympäristöä kuvaavia pisteitä heikentävät menetelmien toimivuutta (Barbet-Massin ym. 2012).

Havaintoihin liittyy niiden luonteen, laadun ja luotettavuuden lisäksi muitakin havaintoaineistoon liittyviä potentiaalisia ongelmia, kuten otoskoko, havaintojen spatiaalinen resoluutio – esimerkiksi havaintojen rasterikoko – ja spatiaalinen autokorrelaatio (Franklin 2009). Suuri otoskoko auttaa minimoimaan tilastollisiin menetelmiin liittyviä riskejä (Guisan & Zimmermann 2000). Riittävä otoskoko varmistaa, että havaintoja on riittävästi suhteessa käytettyihin muuttujiin, suurta aineistoa voidaan jakaa osiin mallin sovittamista ja evaluaatiota varten, minkä lisäksi yksittäisten mittausvirheiden merkitys pienenee mal- leissa, kun havaintoja on tarpeeksi. Loisel ynnä muiden (2008) mukaan pieni otoskoko

on suurempi uhka kuin aineiston keräämistavasta johtuva vinouma. Osa mallinnusmenetelmistä on herkempiä pienille otoskoille kuin toiset (Hernandez ym. 2006). Jos otoskoko on pieni eikä havaintojen määrää voi kasvattaa, tulee käyttää sellaisia menetelmiä, joiden suorituskyky ei heikkene merkittävästi otoskoon pienentyessä (Elith ym. 2006; Wisz ym. 2008b). Riittävä havaintojen määrä on lajikohtaista ja riippuu esimerkiksi tutkimuksen spatiaalisesta skaalasta (Elith ym. 2006), mutta erityisen pieniä otoskokoja, kuten alle 30 havaintoa, tulee välttää (Wisiz ym. 2008b). Tärkeää kuitenkin on, että havainnot kattavat riittävästi erilaisia ympäristöjä tarkastellulta levinneisyysalueelta (Soberón 2007).

Tutkimuksen spatiaalisen mittakaavan ja resoluution valinta vaikuttaa levinneisyysmallinnukseen merkittävästi. Kun tutkimusalue on määritelty, ensimmäinen tärkeä päätös on määrittellä, kuinka suuren alueen yksittäinen lajin yksilöistä tehty esiintymishavainto kattaa eli kuinka suureksi havaintojen rasterikoko määritetään (Guisan ym. 2017: 135). Valintaa tulee suhteuttaa tutkimusalueen kokoon. Mitä tarkempi resoluutio on eli mitä pienempi on rasterikoko, sitä enemmän havaintoja saadaan ja sitä tarkemmin ne paikantuvat ympäristöönsä. Jos tutkimusalue on hyvin suuri, voi olla perusteltua käyttää suurempaa rasterikokoa esimerkiksi tietoteknisistä syistä, ja koska lajin esiintymistä voi suuressa mittakaavassa ohjata tekijät, jotka eivät ilmene tarkalla resoluutiolla (Guisan ym. 2007; Austin & Van Niel 2011). Spatiaalinen resoluutio tulee valita tutkimuskohtaisesti tai vertailla tutkimuksen aikana erilaisten resoluutioiden vaikutusta tuloksiin (Luoto ym. 2007). Suuri resoluutio on enimmäkseen toivottavaa, koska se mahdollistaa suuremman vaihtelun habitaattien välillä esimerkiksi toisilleen läheisillä alueilla ja näin todennäköisesti levinneisyyttä paremmin selittävät mallit (Guisan ym. 2007).

Spatiaalinen autokorrelaatio on yleinen ongelma mallinnettaessa ilmiöitä maantieteellisessä tilassa (Legendre 1993). Levinneisyysmallinnuksessa spatiaalinen autokorrelaatio on läsnä lajihavainnoissa ja ympäristömuuttujissa tarkoittaen, että havainnot ovat keskittyneet samoihin ympäristöihin ja että muuttujan arvot ovat sitä samanlaisempia, mitä lähempänä ne esiintyvät toisiaan (Miller ym. 2019). Vaikka tämä on luonnollista ekologisissa ja maantieteellisissä ilmiöissä, selittämättömänä vaihteluna lajin esiintymisissä ilmevä spatiaalinen autokorrelaatio voi heikentää tulosten luotettavuutta ja yleistettävyyttä erityisesti käytettäessä monimutkaisia algoritmeja (Merow ym. 2014). Havaintojen spatiaalisen autokorrelaation ongelmaa voidaan pienentää esimerkiksi valitsemalla mallinnukseen mukaan vain havaintoja, jotka ovat tietyn matkan päässä toisistaan, mikä voi kuitenkin vähentää mallin sovittamiseen käytettävien havaintojen määrää (Dormann 2007).

Tavallinen keino tarkastella spatiaalisen autokorrelaation läsnäoloa mallinnuksessa on mallin residuaalien eli jäännösarvojen tarkastelu mallin sovittamisen jälkeen (Legendre 1993).

Ympäristömuuttujat kuvaavat habitaatin ominaisuuksia

Levinneisyysmallinnuksessa lajin esiintymistä selittävien ympäristömuuttujien tulee olla karttamuodossa eli rastereina ja vektoreina. Muuttujat ovat tyypillisesti rasterimuodossa, mikä käytännössä tarkoittaa, että alue on jaettu samankokoisiin soluihin, joista jokaisella on muuttujan vaihtelua tai luokkaa kuvaava tunnusarvo ja joihin lajihavaintoja sitten verrataan (Franklin 2009: 77). Muuttujia kuvaavilla rastereilla tulee olla keskenään sama spatiaalinen resoluutio ja niiden tulee kattaa koko se alue, jota lajin levinneisyyden mallinnuksessa käytetään (McPherson ym. 2006; Guisan ym. 2017: 139).

Valmista GIS- tai vastaavaa aineistoa on saatavilla useista lähteistä, mutta käyttäjän tulee pohtia niiden käyttökelpoisuutta erikseen kullekin lajille (Franklin 2009: 77). Muuttujien luontiin ja prosessointiin joudutaankin usein panostamaan, jotta ne olisivat tutkimuksen tavoitteiden ja kohdelajien kannalta merkityksellisiä ja mahdollistaisivat eri menetelmien käytön ja vertailun. Muuttujien laatuun liittyy aina epävarmuutta ja niihin liittyvien virheiden löytäminen aineistosta voi olla mahdotonta (Guisan ym. 2017). Havaintojen luotettavuudella tai onnistuneella tausta-aineiston otantaan käytettävän alueen valinnalla ei ole merkitystä, jos emme kykene löytämään selittäviä tekijöitä lajin esiintymisen taustalla (Austin 2002) ja siirtämään näitä tekijöitä mallinnuksessa käytettäväksi muuttujiksi (Guisan & Zimmermann 2000; Austin 2007).

Austin (2007) painottaa ekologisen teorian merkitystä muuttujien valinnassa. Muuttujat voivat perustua koeasetelmissa tai luonnossa tehtyihin havaintoihin tai teoreettisiin hypoteeseihin. Malleissa käytettyjen muuttujien tulisi olla kausaalisia eli kuvata suoraan lajin levinneisyyteen positiivisesti tai negatiivisesti vaikuttavia tekijöitä sen sijaan, että etsittäisiin vain tilastollisia malleja parantavia muuttujia, joilla ei välttämättä ole todellista merkitystä tutkittavalle lajille (Austin 2002, 2007). Lopulliseen malliin valikoituvat muuttujat, jotka parhaiten ilmentävät lajin ja ympäristön suhdetta noudattaen samalla menetelmien muuttujille asettamia vaatimuksia (Guisan & Zimmermann 2000). Selittävänä muuttujana on käytetty lähinnä ympäristötekijöitä, jotka vaikuttavat lajiin mutta joihin laji itse ei vaikuta: tällaisia ovat esimerkiksi globaalit ilmastotekijät (Jackson & Overpeck 2000; Chase & Leibold 2003). Jos käytämme muuttujaa, jonka arvojen ilmenemiseen ympäristössä laji vaikuttaa, kyseisen tekijän vaikutus lajiin voi peittyä (Peterson ym. 2011).

Joudumme kuitenkin usein tilanteeseen, jossa tärkeitä ympäristötekijöitä kuvaavia muuttujia ei ole saatavilla (Barry & Elith 2006) tai joudumme hyödyntämään muuttujia, jotka eivät varsinaisesti vaikuta suoraan lajin fysiologiaan vaan välillisesti lajin esiintymistä ohjaavaan tekijään (Mod ym. 2016). Epäsuorien muuttujien käytännöllisyys mallinnuksessa riippuu usein tutkimusalueen koosta ja spatiaalisesta resoluutiosta (Austin 2002; Austin & Van Niel 2011). Käytettäessä epäsuoria muuttujia on oltava huolellinen pohdittaessa, minkä verran tuloksia voi yleistää tai siirtää muihin ympäristöihin (Mod ym. 2016).

Jokaisen lajin esiintymiseen vaikuttavan tekijän käyttö muuttujina on mahdotonta ja parasta olisikin yksinkertaisesti etsiä sopiva yhdistelmä tärkeimpiä muuttujia, joilla kaikilla on vaikutusta lajin esiintymiseen (Barry & Elith 2006). Joissain tapauksissa on turvaututtava epäsuorasti vaikuttaviin muuttujiin huolimatta niihin liittyvistä riskeistä (Mod ym. 2016). Mallinnuksessa käytettävät muuttujat tulee valita huomioiden lajin ekologiset vaatimukset, ympäristötekijöiden siirtämisessä muuttujiksi ilmenevät rajoitteet sekä tutkimuksen spatiaalinen mittakaava eli se, kuinka suurella alueella lajin esiintymistä tarkastellaan ja mikä on aineiston havaintojen ja muuttujien resoluutio (Guisan & Zimmermann 2000; Guisan & Thuiller 2005; Soberón 2007).

Kaikki ympäristötekijät eivät vaikuta lajin levinneisyyteen samalla maantieteellisellä mittakaavalla vaan eri tekijöiden rooli vaihtelee siirryttäessä yksittäisistä lajin yksilöistä – joiden esiintymistä ja määrää tarkastellaan paikallistasolla – kohti koko lajin globaalia levinneisyysaluetta (Mackey & Lindenmayer 2001; Guisan & Thuiller 2005). Ilmasto esimerkiksi on merkittävä tekijä lajin esiintymiselle sekä globaalilla tasolla että paikallisesti, mutta maaperä tai kasvillisuuden jakautuminen vaihtelevat alueittain ja paikallisessa mittakaavassa enemmän tarjoten yksilöille ja populaatioille vaihtoehtoisia ympäristöjä (Chase & Leibold 2003). Tämän vuoksi mallinuksissa hyödynnettäviä muuttujia valitessa on tärkeää myös pohtia niiden merkitystä lajin esiintymiselle siinä spatiaalisessa mittakaavassa, jolla levinneisyyttä kulloinkin ollaan tarkastelemassa (Soberón 2007).

Havaintojen, ja niiden jakautumista ympäristöön selittävien muuttujien resoluution on järkevää asettaa vastaamaan toisiaan suurimpien mittakaavan poikkeamisesta aiheutuvien ongelmien välttämiseksi (McPherson ym. 2006). Jos havaintojen resoluutio on suuri eli havaintojen ollessa määritettynä pienellä rasterikoolla, matalalla resoluutiolla mitatut muuttujat eivät kykene erottamaan pienellä alueella tapahtuvasta habitaattien vaihtelusta johtuvia eroja lajin esiintymisessä johtaen näin muodostettavien mallien epätarkkuuteen. Ympäristötekijän matala resoluutio voi olla muutenkin ongelmallista, sillä

ympäristömuuttujasta laskettu tunnusluku, kuten keskiarvo, ei välttämättä kuvaa riittävän hyvin ympäristötekijän ja lajin välistä suhdetta ympäristön vaihdellessa voimakkaasti hilan sisällä (Guisan ym. 2007; Guisan ym. 2017: 138, 140). Ympäristön ja lajin suhteen mallintamista pienellä resoluutiolla heikentää myös epävarmuus siitä, kuinka suuren osan raste-rimuotoisista tutkimusruuduista laji asuttaa; lajin kannalta merkitykselliset habitaatit voivat sijaita verrattain pienellä alueella (McPherson ym. 2006).

Maantieteellisen alueen kasvaessa suurella resoluutiolla mitattujen muuttujien vaikutus ei välttämättä paranna mallien ennusteita. Sama koskee myös pienemmissä mittakaavassa vaikuttavia ympäristötekijöitä ja niiden hyödyntämistä laajojen alueiden tarkasteluissa (Guisan ym. 2007). Alueen kasvaessa kasvaa myös ympäristömuuttujien vaihteluväli: tällöin on todennäköisempää löytää ne muuttujan raja-arvot, jotka ovat lajin esiintymisen kannalta tärkeitä (Kadmon ym. 2003). Matala resoluutio ja mallintaminen suurella alueella eivät automaattisesti johda huonompiin tuloksiin tai malleihin kuin suuren resoluution käyttö paikallisella tutkimusalueella (Guisan ym. 2007; Luoto ym. 2007) vaan resoluution ja spatiaalisen skaalan valinta riippuu tutkimuksen tavoitteista, kohdelajista ja siihen liittyvistä muuttujista sekä siitä, mihin tuloksia aiotaan hyödyntää (Elith & Leathwick 2009; Guisan & Thuiller 2005). Esimerkiksi paikallisten suojelupäätösten tekeminen tai tulosten ekstrapolointi ympäristöltään poikkeaville alueille ei ole turvallista, jos spatiaaliseen skaalaan liittyviä ongelmia ei ole kyetty ratkaisemaan tai kun tehtyjä valintoja, ja niiden perusteella saatuja tuloksia ei pystytä perustelemaan (McPherson ym. 2006; Jarnevich ym. 2015).

Muuttujien valinnassa tulee pohtia myös, miten lajiin vaikuttavat tekijät vaihtelevat ajallisesti ja miten laji reagoi muutoksiin. Emme aina tiedä, vaikuttavatko lajin nykyiseen levinneisyyteen tämän hetkiset olosuhteet vai sittenkin pidemmän aikavälin historialliset olosuhteet. Käytettävän muuttujan avulla tulisi löytää yhteys muuttujan ja havaintojen välillä siinä hetkessä, jossa havainto on tehty (Guisan ym. 2017: 140–141). Tyypillinen, perusteluiltaan itsestään selvä lähtökohta on valita muuttujan arvot läheltä ajanjaksoa, jona lajihavainnot on tehty. Vaihtoehtona tälle on käyttää saman muuttujan arvoja eri ajanjaksoilta (Guisan ym. 2017: 141). Huonon dispersaaliyvyn omaavat tai pitkäikäiset lajit voivat esimerkiksi esiintyä alueella, joka olosuhteiden muutoksesta johtuen ei ole enää suotuisa (Svenning & Skov 2004). Tällöin muuttujan nykyistä tilaa kuvaavien arvojen käyttäminen selittävänä tekijänä voi tuoda virheellisiä tuloksia (Guisan ym. 2017: 141). Nopeasti liikkuvat lajit taas kykenevät reagoimaan ympäristössä tapahtuviin muutoksiin ja etsimään

optimaalisia olosuhteita liikkeessaan ympäristöjen välillä, jolloin niiden olettaisi esiintyvän alueella, jonka abioottinen ja bioottinen ympäristö mahdollistavat esimerkiksi lisääntymisen (Barve ym. 2011).

Levinneisyysmallinnuksessa voidaan käyttää lukuisia eri muuttujaryhmiä- ja tyypppejä, kunhan ne täyttävät mallinnuksen edellyttämät vaatimukset ja ovat tutkittavan lajin kannalta merkityksellisiä (taulukko 1). Erilaiset ympäristötekijän keskiarvoa, vaihteluväliä, ääriarvoja ynnä muita muuttujan määrää, yleisyyttä ja vaihtelua kuvaavat muuttujat ovat malleissa tyypillisiä (Franklin 2009; Guisan ym. 2017). Ilmastomuuttujia käytetään lähes poikkeuksetta jokaisessa levinneisyysmallinnuksessa lukuun ottamatta pienimmän spatiaalisen mittakaavan tutkimuksia (Guisan & Thuiller 2005). Vuosittaisia lämpötila- tai sadantakeskiarvoja toimivampia parametrejä mallinnuksessa ovat lajille tärkeiden vuodenaikojen äärilämpötilat ja keskiarvot sekä erilaiset kasvukauden pituudesta kertovat muuttujat, joiden vaikutus lajeihin on myös ekologisesti perusteltua (Root 1988). Ilmastomuuttujia tukemaan, tai niiden sijaan, voidaan käyttää muuttujia, jotka kuvaavat topografisia tekijöitä, kuten korkeuseroja ja rinteiden kaltevuutta (Austin & Van Niel 2011). Niiden käyttö rajoittuu pääosin paikallisiin tutkimuksiin korkeusvaihtelun muokatessa alueellisia lämpötiloja, sadantaa ja esimerkiksi maaperän kosteutta ja vaikuttaessa näin lajeihin epäsuorasti (Körner 2007).

Erilaisia maanpeitettä tai -käyttöä kuvaavia muuttujia käytetään kuvaamaan lajien habitaattivaatimusten mukaisen ympäristön saatavuutta ja määrää alueittain tai lajiin kohdistuvia häiriötekijöitä, kuten ihmisvaikutusta (Wisz ym. 2008a). Levinneisyysmallinnuksessa hyödynnetään kasvavissa määrin myös kaukokartoitusmenetelmin (kaukokartoitus = *remote sensing*) tuotettuja muuttujia, jotka kuvaavat esimerkiksi kasvillisuutta ja sen tilaa sekä habitaattien laatua: kaukokartoituksen roolin odotetaan kasvavan tulevaisuudessa ja tukevan perinteisiä ympäristömuuttujia lajien levinneisyyden mallintamisessa (He ym. 2015; Kissling ym. 2018). Koska bioottisilla vuorovaikutussuhteilla on merkittävä rooli lajien esiintymisessä, niiden hyödyntäminen muuttujina levinneisyysmallinnuksessa olisi tärkeää (Wisz ym. 2013; Dormann ym. 2018). Muiden lajien vaikutuksen tuominen malleihin luotettavasti voi kuitenkin olla vaikeaa ellei jopa mahdotonta (Kissling ym. 2012). Levinneisyysmallinnuksessa hyödynnettävien bioottisten muuttujien tulisi esimerkiksi olla sellaisia, joihin laji itse ei vaikuta (Anderson 2017) ja esimerkiksi ravintoverkkojen kaltaisten monimutkaisten vuorovaikutussuhteiden roolia on mahdotonta esittää perinteisissä malleissa (Wisz ym. 2013; Dormann ym. 2018). Myös kiistat lajien välisen vuorovaikutuksen

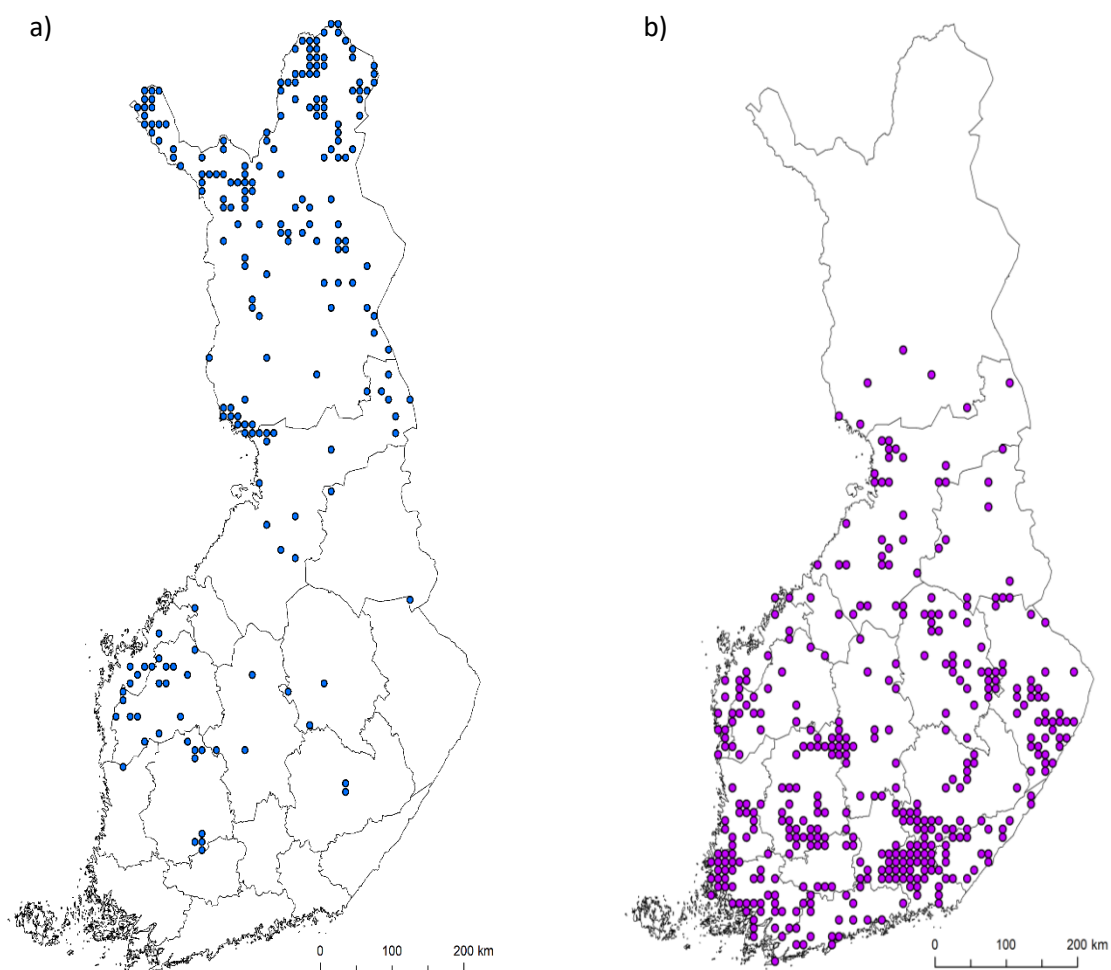
ilmenemisestä eri spatiaalisilla mittakaavoilla ovat vaikuttaneet bioottisten muuttujien käyttöön (Guisan ym. 2017). Vaikka perinteisesti on ajateltu, että bioottiset tekijät vaikuttavat lajien esiintymiseen vain paikallisesti (Chase & Leibold 2003), on sittemmin esitetty todisteita myös sen puolesta, että bioottisten vuorovaikutussuhteiden vaikutus näkyy laajemminkin lajien levinneisyydessä (Wisiz ym. 2013; Araújo & Rosenfeld 2014), jopa globaalissa mittakaavassa (Pigot & Tobias 2012).

Taulukko 1. Levinneisyysmallinnuksessa tyypillisesti käytettyjä ympäristömuuttujia.

Muuttujaryhmä	Ympäristötekijä, esimerkki	Muuttuja	Esimerkki tutkimuksesta
Ilmasto	Lämpötila	Vuosittainen, vuodenaikainen	Virkkala ym. 2005
	Sadanta	lämpötila, sadanta, säteily	Elith ym. 2006
	Säteilyn määrä	(esim. keskiarvo, ääriarvot)	Phillips ym. 2006
Topografia	Korkeus merenpinnasta	Korkeus, korkeusvaihtelu	Kosicki 2017
	Rinteen kaltevuus	Kasvupaikan saama säteily	Title & Bemmels 2018
	Rinteen suunta	Valunta	
Maanpeite ja/tai maankäyttö: habitaatit ja häiriöt	Kasvillisuustyytit	Lajin suosiman habitaatin	Luoto ym. 2007
	Vesistöt, kosteikot	(metsä, kosteikko, niitty,...)	Titeux ym. 2007
	Ihmisvaikutus	määrä tai peittävyys (esim. keskiarvo) Etäisyys habitattiin, häiriötekijään	Shirley ym. 2013
Kaukokartoitus	Kasvillisuus	Kasvillisuusindeksit (mm.	Cord & Rödder 2011
	Ekosysteemin tuottavuus	NDVI)	Pottier ym. 2014
	Maaperän ominaisuudet	Maaperäindeksit	Deblauwe ym. 2016
Bioottiset tekijät	Kilpailevien-, saalislajien, muiden	Muiden lajien esiintymisha- vainnot tutkimusruuduissa	Heikkinen ym. 2007 Meier ym. 2010
	kohdelajiin vaikuttavien lajien levinneisyys	Muiden lajien määrä tutkimusruuduissa	Guisan & Rahbek 2011
	Lajirunsaus	Habitaatin tuottavuus (esim.	
	Habitaatin ”kantokyky”	kaukokartoituksen muuttujat)	

Tutkimusalue

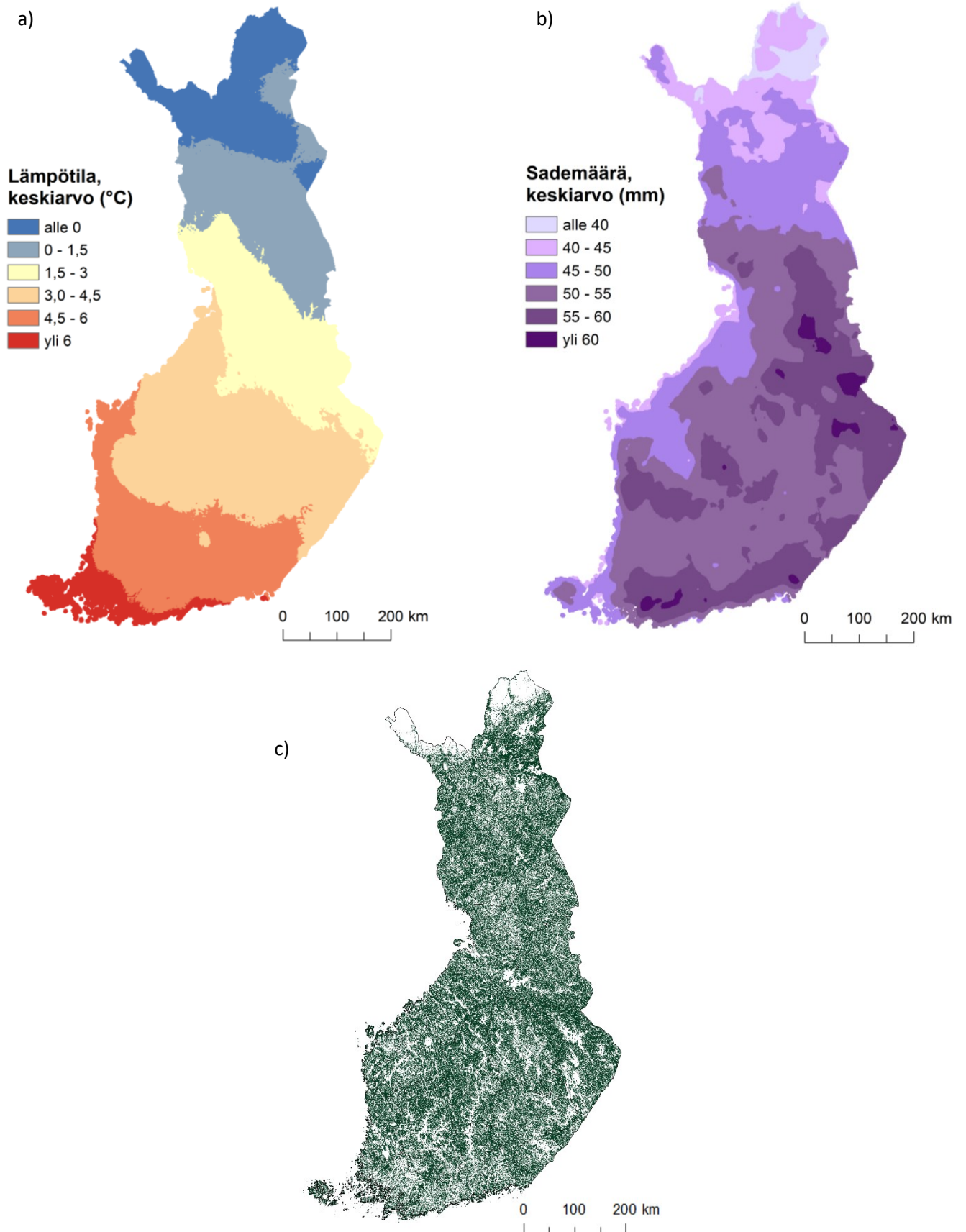
Tutkimusalueena on pohjoisessa Euroopassa sijaitseva Suomi. Mallinnuksessa käytetään ainoastaan Suomessa tehtyjä lajihavaintoja ja Suomen ympäristöä kuvaavia muuttujia, minkä lisäksi itse ennusteet tehdään koko Suomen alueelle huolimatta painotuksista lajin esiintymisessä tutkimusalueen eri osiin (kuva 3). Tutkimuksen painottuminen lajin Suomen levinneisyysalueeseen tarjoaa mahdollisuuden esimerkiksi tutkia lajin alueelliseen esiintymiseen vaikuttavien tekijöiden merkitystä kyseisille lajeille (Guisan ym. 2017: 148). Suomi sijaitsee useiden päiväpetolintujen levinneisyyden pohjoisrajalla (The IUCN Red List... 2020), jolla yksilöiden esiintymiseen voivat vaikuttaa eri tekijät kuin muualla lajin globaalin levinneisyysalueen sisällä (Sexton ym. 2009).



Kuva 3. A) Piekanan ja b) mehiläishaukan esiintymishavainnot (Suomen Lajitietokeskus 2020, 10 km x 10 km -ruudun keskipiste).

Suomi kuuluu Köppenin ilmastoluokituksessa kylmätalviseen lumi- ja metsäilmastoalueeseen, jolle tyypillistä ovat esimerkiksi kylmät talvet ja ympärivuotiset sateet. Suomen ilmastossa on sekä manner- että meri-ilmaston piirteitä ja lämpötilojen vaihtelu vuodenaikojen välillä on voimakasta kautta maan. Lounais- ja Etelä-Suomessa talvet voivat olla huomattavasti pohjoista leudompia ja talven sateet tulevat kasvavissa määrin vetenä lumen sijaan. Lämpimin kuukausi on yleisesti heinäkuu, kun taas tammi- ja helmikuu ovat vuoden kylmimpiä kuukausia keskilämpötilaltaan. Sateisinta on heinä-elokuussa ja kuivinta keväällä sekä keskitalvella (Kersalo & Pirinen 2009: 11–13). Vuosien 2000 ja 2013 välillä lämpötilan vuotuinen keskiarvo vaihteli noin $-2\text{ }^{\circ}\text{C}$ ja $7\text{ }^{\circ}\text{C}$ välillä koko Suomen keskiarvon ollessa noin $2,7\text{ }^{\circ}\text{C}$ (kuva 4a). Sademäärän kuukausittaisen keskiarvon vaihtelu samalla ajanjaksolla oli noin 35 millimetrin ja noin 65 millimetrin välillä koko Suomen keskiarvon ollessa noin 52 millimetriä kuukaudessa (kuva 4b). Edellä mainitut tunnusluvut ovat peräisin Ilmatieteen laitoksen paikkatietoaineistoista johdetuista muuttujista, joita käytettiin mallinnuksessa.

Suomi kuuluu boreaaliseen kasvillisuusvyöhykkeeseen, joka voidaan jakaa edelleen viiteen alavyöhykkeeseen: hemi-, etelä-, keski- ja pohjoisboreaaliseen sekä hemiarktiseen vyöhykkeeseen (Virkkala ym. 2000). Suomen vaihteleva ilmasto tarjoaa monipuolisia kasvuympäristöjä erilaisille boreaalisen vyöhykkeen kasvilajeille, mutta pääosa esimerkiksi Suomen metsistä koostuu havupuista eli männystä ja kuusesta sekä havu- ja lehtipuiden muodostamista sekametsistä. Lehtipuiden osuus puustosta pienenee siirryttäessä kohti pohjoista. Pohjoisimmassa Suomessa kuusimetsät vaihtuvat tuntureiden niukkapuustoihin alueisiin, jossa kasvillisuus, eritoten puuston ja metsien osalta, on selvästi vähäisempää kuin muualla Suomessa (Virkkala ym. 2000). Metsät kattavatkin suuren osan Suomen luonnosta (kuva 4c). Asutusta on harvassa ja se keskittyy erityisesti Etelä-Suomeen ja erilaisiin keskuksiin eri puolilla Suomea. Suuret alueet Lapissa ja Itä-Suomessa ovat erittäin harvaan asuttuja. Suomen luonto on tunnettu soiden runsaasta määrästä, minkä ovat mahdollistaneet kohtalaiset sateet ja viileä ilmasto, joka vähentää veden haihduntaa (Kersalo & Pirinen 2009). Huolimatta metsien ja muiden luonnonympäristöjen runsaasta määrästä metsien rakenne on muuttunut yksipuolisemmaksi metsätalouden seurauksena ja linnuston suosimat, täysin luonnontilaiset metsät ovat harvassa painottuen haastavien kasvuolosuhteiden Lappiin (Laaksonen & Lehikoinen 2013).



Kuva 4. Vuosien 2000–2013 a) lämpötilan (Ilmatieteen laitos 2020a, 2020c) ja b) sademäärän keskiarvo (mm/kk, Ilmatieteen laitos 2020b, 2020d) Suomessa ja c) Suomen metsien jakautuminen vuonna 2006 (vihreä = metsä, SYKE 2010).

Aineisto ja menetelmät

Tutkimusaineisto

Lajihavaintoina toimivat piekanan ($n = 204$) ja mehiläishaukan ($n = 401$) vuosien 2000–2015 välillä tehdyt kansalaishavainnot ja suomalaisen rengastusrekisterin havaintoaineisto (Suomen Lajitietokeskus/FinBIF 2020). Kyseisten lajien uhanalaisuuden vuoksi yksilöiden tarkkaa sijaintia ei ollut saatavilla tähän tutkielmaan. Kaikille julkinen yksilöiden sijainti esitetään sen sijaan sadan neliökilometrin ($10 \text{ km} \times 10 \text{ km}$) suuruisissa ruuduissa siten, että ilmoitetut koordinaatit ovat ruudun keskipiste. Havaintoaineisto on niin sanottua *presence-only* -muotoista dataa, sillä se sisältää vain tiedon ruuduista, joissa lajia tavataan – näiden ulkopuolella olevista ruuduista laji joko puuttuu tai sen läsnäoloa ei ole havaittu. Mallinuksessa yhtä havaintopistettä kohden säilytettiin ainoastaan yksi havainto, vaikka samasta ruudusta olisi ollut useampi esiintymishavainto. Havainnoiksi kelpuutettiin vain lajien lisääntymisaikana tai arvioituilla lisääntymispaikoilla tehdyt havainnot, jotta ne kuvaisivat mahdollisimman hyvin lajille soveltuvaa ympäristöä. Havaintojen luotettavuuden arviointi on ollut havainnoitsijoiden oman tulkinnan varassa, mutta lopullisesta aineistosta on pyritty poistamaan epävarmat havainnot.

Rasteriaineistona tallennetut ympäristömuuttujat (taulukko 2) kuvaavat Suomen lämpötilaa ja sadantaa sekä erilaisia maanpeitetyyppejä tai maankäyttöä. Kaikki ilmasto- ja sadantamuuttujat ovat vuosina 2000–2013 mitattuja lämpötilojen ja sademäärien keskiarvoja muuttujan nimen mukaisella ajanjaksolla. Esimerkiksi muuttujaa ”huhti-toukokuun lämpötila” kuvaava arvo on huhti- ja toukokuun keskimääräinen lämpötila. Alkuperäisten ilmastorastereiden tuottajana on toiminut Ilmatieteen laitos (2020a, 2020b, 2020c, 2020d). Loput muuttujat kuuluvat toiseen ryhmään eli maanpeitemuuttujiin. Ne kuvaavat esimerkiksi erityyppisten luonnonympäristöjen, kuten metsien tai soiden, tai rakennetun ympäristön prosentuaalista osuutta havaintoyksikön eli rasterin solun pinta-alasta pyrkien ilmentämään erilaisten habitaattien saatavuutta. Alkuperäinen aineisto on Suomen ympäristökeskuksen tuottama, Euroopan ympäristöviraston CLC2006-projektin mukaisia Suomen CORINE-maanpeitetyyppejä ja maankäyttömuotoja vuonna 2006 kuvaava aineisto, joka perustuu pääosin satelliittikuvatulkintaan (SYKE 2010). CORINE-aineiston on osoitettu sopivan levinneisyysmallinnukseen Suomessa, kun kohteena olivat soiden linnusto (Virkkala ym. 2005). Alkuperäiset CORINE-luokat luokiteltiin uudelleen vastaamaan paremmin tutkimuksen tarpeita.

Taulukko 2. Ympäristömuuttujat, joiden avulla lajien levinneisyyttä pyritään ennustamaan. Maanpeitemuuttujien arvot kuvaavat luokan **peittävyyttä** (%) havaintoruudussa, lämpötila on **celsiusasteina** ja sademäärä **millimetreinä**.

Aineisto	Muuttuja	Kuvaus (maanpeite-/ maankäyttömuuttujat)	Corine- luokat (taso 4)	Keskiarvo, (min – max)
SYKE, CO- RINE 2006	Ihmistoiminta	Rakennettu ympäristö, asutus, ihmisen aiheuttamat häiriötekijät	1–13	0,04 (0 – 0,60)
	Maatalous	Maatalousalueet: pellot, laidunmaat	14–17	0,08 (0 – 0,66)
	Metsä	Lehti-, havu-, sekametsät	18–25	0,54 (0 – 0,85)
	Avoin tai harvapuustoinen luonnonympäristö	Luonnonniityt, harvapuus- toiset alueet, kalliomaat, rantahietikot ja dyynit	26–35	0,15 (0 – 0,96)
	Suot ja kosteikot	Avosuot, sisämaan- ja merenrantakosteikot, turvetuotantoalueet	36–41	0,08 (0 – 0,67)
	Vesistöt	Joet, järvet, meret	42–44	0,11 (0 – 0,94)
Ilma- tieteen laitos 2000- 2013	Lämpötila, huhti- elokuu			10,8 (5,7 – 13,4)
	Lämpötila, huhti- kesäkuu			7,7 (2,6 – 10,5)
	Lämpötila, huhti- toukokuu			5,1 (-0,0 – 8,1)
	Lämpötila, touko- kesäkuu			10,7 (5,4 – 13,3)
	Lämpötila, kesä- elokuu			14,5 (9,4 – 17,0)
	Sademäärä, huhti- elokuu			58,0 (40,9 – 70,8)
	Sademäärä, huhti- kesäkuu			47,4 (34,5 – 60,7)
	Sademäärä, huhti- toukokuu			39,1 (29,11 – 52,9)
	Sademäärä, touko- kesäkuu			55,9 (36,7 – 72,5)
	Sademäärä, kesä- elokuu			71,8 (51,3 – 86,4)

Alkuperäisten ilmastomuuttujien resoluutio oli yksi kilometri ja CORINE-aineiston muuttujien 25 metriä. Levinneisyysmallinnusta varten alkuperäisten aineistojen resoluutio yleistettiin vastaamaan lajihavaintojen resoluutiota (10 km x 10 km). Ympäristömuuttujat on pyritty ajoittamaan samalle ajanjaksolle kuin esiintymishavainnot (2000-2015), jotta ne kuvaisivat mahdollisimman hyvin lajien ympäristöä havaintojen ajankohtina. Muuttujien valmistelu suoritettiin Esri ArcMap -ohjelmistolla (2019, versio 10.7.1). Muuttujat valittiin kuvaamaan ympäristötekijöitä, jotka voivat vaikuttaa piekanan ja mehiläishaukan sekä

muiden päiväpetolintujen levinneisyyteen Suomen laajuisella tutkimusalueella. Kevään ja alkukesän lämpötilojen ja sadannan keskiarvot kuvaavat kevätmuuton ja pesinnän aloituksen aikaisia ilmasto-olosuhteita, joiden oletetaan vaikuttavan esimerkiksi lintujen pesimäpaikan valintaan ja näin niiden levinneisyyteen (Heikkinen ym. 2006). Pelkkien kesäkuukausien (kesä-elokuu) ja huhti-elokuun ilmastomuuttujien lisäämisessä tarkasteluihin on tarkoituksena selvittää, ovatko nimenomaan muuton ja pesinnän aloituksen aikaiset ilmasto-olosuhteet tärkeimmät lajien kannalta. Mehiläishaukan muutto ja pesinnän aloitus tapahtuvat hieman myöhemmin kuin piekanan (Gensbøl ym. 1995), mikä on myös huomioitu muuttujavalinnassa. Maanpeitemuuttajat valittiin kuvaamaan lajien käytössä olevia habitaatteja. CORINE-muuttujien lisääminen ilmastomuuttujien rinnalle on parantanut aiemmin esimerkiksi suolintujen levinneisyysmallinnusta Suomessa (Virkkala ym. 2005).

Tutkimusvaiheet ja -menetelmät

Tutkittavien lajien varsinainen levinneisyysmallinnus suoritettiin kahdessa vaiheessa. Ensimmäisessä vaiheessa vertailtiin erilaisten levinneisyysmallinnuksen menetelmien ennustuskkyä. Tarkastelun avulla menetelmien joukosta pyrittiin rajaamaan parhaiten päiväpetolintujen levinneisyysmallinnukseen soveltuvat menetelmät. Tutkielmassa raportoituna ovat vain parhaita ennusteita tuottaneet menetelmät, kummallekin lajille erikseen. Valittujen menetelmien katsottiin tuottaneen tasaisimpia ja parhaita ennusteita ja erottuneen edukseen lukuisten menetelmien joukosta. Toisessa vaiheessa yksittäisen mallin sovittaminen, evaluointi ja lopullinen ennuste lajien levinneisyydestä suoritettiin molemmille lajeille menetelmällä, jonka havaittiin suoriutuneen hyvin ensimmäisen vaiheen mallinnuksissa. Kyseinen menetelmä oli GBM (*Generalized Boosted Model*, Greenwell ym. 2019). Toisen vaiheen mallien avulla kyettiin luomaan lajien levinneisyyden ennustekartat ja arvioimaan mallien suorituskykyä ja luotettavuutta *presence only* -aineistoon soveltuvilla arviointikriteereillä, joiden käyttö ensimmäisessä vaiheessa ei ollut mahdollista tai mielekäästä. Kaikki mallinnukset suoritettiin R-ympäristössä (R Core Team 2016).

Levinneisyysmallinnuksen tuloksena syntyy ennuste jota voidaan kutsua *presence-only* -aineistolla lajin suhteelliseksi esiintymistodennäköisyydeksi (Elith & Leathwick 2009). Jokaiselle Suomen ruudulle ennustetaan malleissa arvo, joka vaihtelee välillä 0–1: mitä suurempi arvo on sitä todennäköisempää on havaita laji kyseisessä ruudun sisällä luonnossa. Arvon voidaan myös tulkita kertovan habitaatin sopivuudesta lajin habitaattivaatimusten näkökulmasta suhteessa muihin habitaatteihin (Guisan ym. 2017). On

kuitenkin muistettava, että mallit pyrkivät vain yksinkertaistamaan ja heijastelemaan todellisuutta eivätkä ole siinäkään aina tarkkoja: levinneisyys on ilmiönä sen verran monimutkainen, että ennusteiltaan erinomainenkin levinneisyysmalli voi vain arvioida habitaattien suhteellista sopivuutta tarkastelun kohteena olevalla maantieteellisellä alueella (Franklin 2009).

Levinneisyysmallinnuksen menetelmien ennustuskykyä tulee tarkastella aineistokohtaisesti

Ensimmäinen vaihe, menetelmien vertailu, suoritettiin BIOMOD2-paketin (Thuiller ym. 2020) avulla: se mahdollistaa 11 levinneisyysmallinnuksessa hyödynnetyn menetelmän tai algoritmin yksinkertaisen vertailun tarjoten käyttäjälle mahdollisuuden esimerkiksi valita mallien sovittamisessa käytetyt menetelmäkohtaiset asetukset ja vaihtoehtoja mallien arviointiin. BIOMOD2:n menetelmiin kuuluu sekä perinteisempiä regressiomenetelmiä, kuten GLM (*Generalized Linear Model*) ja GAM (*Generalized Additive Model*), yksinkertaisia, nimenomaan levinneisyysmallinnukseen kehitettyjä menetelmiä, kuten SRE (*Surface Range Envelope*), monipuolisia, tietokoneoppimisen parissa kehitettyjä menetelmiä, kuten ANN (*Artificial Neural Networks*), RF (*Random Forest*) ja GBM sekä pelkkiä esiintymishavaintoja hyödyntävien tutkimusten parissa suosittu MAXENT (*Maximum Entropy Modeling*). BIOMOD2 helpottaa ja ennen kaikkea nopeuttaa menetelmien vertailua mahdollistamalla heikosti suoriutuvien algoritmien rajaamisen vaihtoehtojen ulkopuolelle käyttäjän valitsemien arviointikriteerien perusteella.

Menetelmien vertailua varten oli ensin luotava tausta-aineisto yksilöhavaintojen tueksi, jotta menetelmät kykenivät vertaamaan havaintoja ympäristömuuttujiin ja sovittamaan näin mallit. Tämä suoritettiin BIOMOD2:n sisäisillä toiminnoilla siten, että tausta-aineisto valikoitui satunnaisesti ympäristörastereiden noin 3300 ruudun joukosta kuitenkin niin, että yksilöhavaintojen ruudut eivät voineet valikoitua mukaan (BIOMOD2-asetus: `pa.strategy = disk, dist: 10 000`). Näiden ruutujen jättäminen satunnaisotannan ulkopuolelle paransi ennusteita kaikilla menetelmillä, mikä oli odotettavaa aiempien tutkimusten perusteella (Barbet-Massin ym. 2012). Tausta-aineiston määrän valinta pyrittiin tekemään niin, että ennusteet olivat mahdollisimman hyviä ilman, että yksittäisten mallien ennusteiden välinen vaihtelu kasvaisi liikaa. Koska tausta-aineiston jakautuminen ympäristössä voi vaihdella runsaasti ja vaikuttaa näin mallin ennusteisiin (Phillips ym. 2009), on menetelmien suoriutumista parempi tarkastella sovittamalla useita malleja erilaisilla tausta-aineistoilla.

BIOMOD2-menetelmien joukosta karsittiin heikoimmin suoriutuneet menetelmät käyttämälle yksinkertaisia yhden ja kahden muuttujan malleja. Piekanan levinneisyyttä ennustamaan valikoituivat tämän jälkeen GAM, RF ja GBM. Mehiläishaukalle parhaiten suoriutuneet menetelmät olivat GLM, GAM, MAXENT (maxnet-algoritmi), RF, ja GBM. Menetelmien asetuksina käytettiin pääosin BIOMOD2:n määrittämiä oletusasetuksia, mutta GLM:n ja GBM:n kohdalla asetuksia muokattiin hieman esitarkasteluiden jälkeen (taulukko 2). Lopullisissa vertailuissa käytettiinkin **kymmentä** erilaista satunnaisotosta, jotta tausta-aineiston jakauman vaihtelun aiheuttama vaihtelu mallien ennusteissa paljastuisi. Tausta-aineiston määräksi valikoitui GAM-, GLM- ja MAXENT-menetelmien malleihin 500 kappaletta ja RF- ja GBM-menetelmien malleihin 1500 kappaletta. Nämä määrät tuottivat hyviä tuloksia ja vastasivat Barbet-Massinin ynnä muiden (2012) aikaisempia huomioita tausta-aineiston sopivasta määrästä kyseisillä menetelmillä. Mallien evaluaatio suoritettiin jakamalla yksittäistä mallia varten aineisto siten, että malli sovitettiin käyttäen 70 % havainto- ja tausta-aineistosta ja loppu 30 % jätettiin mallin evaluointiin. Koska yksittäisen mallin ennusteet voivat vaihdella sovittamiseen ja evaluaatioon käytettyjen osa-aineistojen ja tausta-aineiston valikoitumisen myötä, mallien sovittaminen ja evaluointi suoritettiin 20 kertaa. Tämä prosessi suoritettiin jokaiselle kaikki havainnot ja satunnaisen tausta-aineiston sisältävälle otokselle. Yhteensä malleja siis sovitettiin yhtä menetelmää kohden 200 kappaletta (paitsi mehiläishaukan MAXENT, 198 kpl, johtuen tuntemattomasta virheestä mallinnuksessa).

Taulukko 3. Ensimmäisessä vaiheessa tarkastellut menetelmät.

Menetelmä	Esimerkki levinneisyysmallinnuksessa	Asetukset, BIOMOD2
GLM (Generalized Linear Model)	Guisan & Zimmermann 2000	<i>type = quadratic, interaction level = 1</i>
GAM (Generalized Additive Model)	Guisan ym. 2002	BIOMOD2-oletusasetukset
MAXENT (Maximum Entropy Modeling)	Phillips ym. 2006	BIOMOD2-oletusasetukset
RF (Random Forest)	Cutler ym. 2007	BIOMOD2-oletusasetukset
GBM (Generalized Boosted Model)	Elith ym. 2008	<i>n.trees = 5 000, interaction.depth = 5</i>

Levinneisyysmallinnuksen piirissä mallien hyvyttä on arvioitu lukuisin eri menetelmin eikä yhtenäistä näkemystä yleispätevistä evaluaatiokriteereistä ole löydetty (Guisan ym. 2017). Tyypillisimmin evaluaatiokriteerinä on käytetty mallien erottelukykä (*discrimination ability*), joka voidaan esittää yksittäisinä ennustuskykyä mittaavina arvoina (Jiménez-Valverde & Lobo 2007; Franklin 2009; Guisan ym. 2017: luku 8). Erottelukykyä käytettiin mallien hyvyyden mittarina myös tässä tutkielmassa. Se kertoo, kuinka hyvin malli kykenee ennusteissaan erottelemaan lajin esiintymisruudut niistä, joissa lajia ei esiinny (Guisan & Zimmermann 2000). Jotta mallit kykenevät muuttamaan nollan ja yhden välillä vaihtelevat jatkuvat ennusteet binäärisiksi ennusteiksi (1 = laji esiintyy; 0 = laji ei esiinny), tulee ennusteiden jaottelulle määritellä kriteerit, joita noudattamalla nämä binääriset ennusteet noudattaisivat mahdollisimman hyvin aineiston ilmentämää jakaumaa lajin esiintymisessä ja toivottavasti myös lajin todellista esiintymistä luonnossa (Jiménez-Valverde & Lobo 2007). Tämä tapahtuu esimerkiksi hyödyntämällä kynnsarvoista riippuvaisia tai riippumattomia ennusteiden arviointimenetelmiä (Jiménez-Valverde 2012; Liu 2016). Erottelukykyä arvioitaessa tarkastellaan, mikä osa mallin johtamista binäärisistä ennusteista on yhteneväisiä aineiston havaintojen kanssa ja kuinka suuri osa ennusteista ja havainnoista poikkeaa toisistaan eli kuinka hyvin malli ennustaa lajin levinneisyyttä ja luokittelee nämä jatkuvat ennusteet binäärisiksi ennusteiksi (Guisan ym. 2017: 252–253).

Ensimmäisessä tutkimusvaiheessa evaluaatiokriteereinä käytettiin ainoastaan mallien erottelukykä mittaavia TSS- (Allouche ym. 2006, *True Skill Statistic*) ja ROC:n AUC-arvoja (Fielding & Bell 1997, *Area Under Curve of Receiver-Operating Characteristic*). Myös TSS:n muodostavia spesifisyyttä (*specificity*) ja sensitiivisyyttä (*sensitivity*) on tarkasteltu erikseen TSS:n ollessa maksimiarvossa (maxTSS/maxSSS, Liu ym. 2016). Spesifisyys kertoo, kuinka suuri osa lajin poissaoloista (0) on ennustettu oikein sensitiivisyyden mitatessa samaa lajin esiintymisille (1) ruuduissa (Franklin 2009: 215). Sensitiivisyyden arvoa voitaneen pitää hyvänä mittarina tässä tutkielmassa, sillä se tarkastelee vain mallin kykyä ennustaa ruudut, joissa lajia on tavattu (Franklin 2009: 215). Menetelmien ennustuskykyä vertaillaan TSS- ja AUC-arvojen keskiarvojen ja arvojen hajonnan avulla – raportoidut arvot ovat luonnollisesti **evaluaatio-otosten arvoja**.

TSS- ja AUC-arvoja on hyödynnetty levinneisyysmallinnuksessa sekä *presence-absence*- että *presence-only* -aineistoilla (Guisan ym. 2017). Tausta-aineiston käyttäminen malleissa lajin poissaolon sijaan tarkoittaa kuitenkin sitä, että TSS- tai AUC-arvot eivät ole tarkkoja arvioita mallin kyvystä erottaa lajin levinneisyyden jakautumista luonnossa

(Phillips & Elith 2010; Guisan ym. 2017: 265). Ne voivat periaatteessa mitata tarkasti vain sen, kuinka hyvin mallin ennusteet korreloivat ruuduissa, joissa lajin on havaittu esiintyvän. AUC-arvo on kohdannut myös huomattavaa kritiikkiä ja sen käytössä on painotettu kiinnittämään huomiota esimerkiksi tutkimuksen tavoitteisiin ja aineiston luonteeseen (Barbet-Massin ym. 2012; Jiménez-Valverde 2012). BIOMOD2 tarjoaa rajallisen määrän evaluaatiokriteerejä mallien arviointiin. Huomioiden TSS:n ja AUC:n yleisen käytön vastaavia aineistoja hyödyntävissä tutkimuksissa kyseisten evaluaatioarvojen arvioitiin toimivan yhdessä riittävänä mittarina arvion eri menetelmien ennustuskyvystä.

Muuttujat valikoitiin tarkastelemalla niiden ekologista uskottavuutta, vaikutuksen suuntaa ja voimakkuutta lajin esiintymistodennäköisyyteen, muuttujan suhteellista tärkeyttä malleihin ja muuttujien lisäämisen tai poistamisen vaikutusta mallien keskimääräiseen ennustuskyykyyn. Vaikutuksen suuntaa ja voimakkuutta arvioitiin vastekäyrillä, jotka luotiin Elithin ynnä muiden (2005) mukaan keskiarvoistamalla muiden muuttujien arvot. Kaksiulotteisista vastekäyristä pystyi tunnistamaan epäuskottavia suhteita lajin ja ympäristömuuttujan välillä, mutta täysin luotettavina mittareina lajin ja muuttujan suhteesta niitäkään ei tule pitää (Zurell ym. 2012). Muuttujien merkitys malleille arvioitiin BIOMOD2:n *variable importance* -valintaa hyödyntäen. Multikollineaarisuuden välttämiseksi muuttujien keskinäisiä korrelaatioita tarkasteltiin ennen mallinnusta laskemalla Pearsonin korrelaatiokerroin, minkä lisäksi käytettävät muuttujayhdistelmät tarkastettiin korkeiden VIF-arvojen (*variance inflation factor*) varalta. Lämpötilamuuttujat korreloivat voimakkaasti keskenään samoin kuin sadantaa kuvaavat muuttujat (liite 1). Ainoastaan kevään (huhtitoukokuu) ja kesän (kesä-elokuu) sademäärien keskinäinen korrelaatio oli vähäinen (0,352). Muuttujien päällekkäisyyden ja korkeiden korrelaatioiden seurauksena malleihin kelpuutettiin vain yksi lämpötilamuuttuja ja yksi sademuuttuja. Maanpeitettä kuvaavista muuttujista ainoastaan soiden ja kosteikoiden määrä korreloi melko voimakkaasti lämpötilamuuttujien kanssa.

Muuttujien määrä pyrittiin pitämään minimissä, koska yksittäisten mallien sovittamisen tarkastelu oli ja suuri määrä selittäviä muuttujia voisi altistaa mallit voimakkaammalle ylisovittamiselle (Holt ym. 2009). Käytetyt menetelmät kykenevät rajoittamaan mallien monimutkaisuutta muun muassa rajoittamalla muuttujien määrää, kuten regressio-menetelmien askeltavat (*stepwise*) menetit tai GBM:n ja RF:n *boosting*- ja *bagging*-metodeihin liittyvien vaiheiden avulla (Elith ym. 2008; Guisan ym. 2017). Vaikka mallinnusten tavoitteena oli löytää mahdollisimman hyviä ennusteita tutkimusaineistolle tuottavia

malleja, ylisovittamisella katsottiin oleva epätoivottu vaikutus vertailuun. Ylisovittaminen lisää mallien välistä ennustuskyvyn vaihtelua ja ylisovitettu malli voi myös arvioida virheellisesti habitaattien suhteellista sopivuutta lajeille vain sen perusteella, kuinka havainnot ovat painottuneet käytetyssä aineistossa, joka sekin sisältää todennäköisiä virheitä esimerkiksi puuttuvien lajihavaintojen myötä (mm. Merow ym. 2014).

Levinneisyysmallinnuksen ennusteita tulee tarkastella aineistoon sopivilla arviointikriteereillä

Toinen vaihe sisälsi siis yksittäisten mallien sovittamisen GBM-menetelmällä sekä piekalle että mehiläishaukalle. Sovittamisessa ja evaluaatiossa hyödynnettiin ominaisuuksia R-paketeista *gbm* (Greenwell ym. 2019), *dismo* (Hijmans ym. 2017) ja *ecospat* (Di Cola ym. 2020). Mallien sovittaminen yksitellen antoi mahdollisuuden evaluoida erottelukyvyn lisäksi myös mallin kalibrointia, muuttujien välisiä suhteita ja sovittamisessa tehtyjen valintojen merkitystä. GBM on niin sanottuihin BRT-menetelmiin (*Boosted Regression Trees*) perustuva menetelmä, joka yhdistää päätöspuualgoritmit niin kutsuttuun *boosting*-menetelmään, jolloin useiden päätös- tai regressiopuiden ennusteet voidaan yhdistää ennusteiden parantamiseksi pitäen huolta samalla, että malli ei monimutkaistu enempää kuin on tarpeen (Elith ym. 2008). Menetelmään liittyy useita etuja esimerkiksi regressiomenetelmiin verrattuna (Elith ym. 2008; Franklin 2009: 165–169). Se ei aineistolähtöisenä menetelmänä aseta ennakkovaatimuksia aineistolle, on mukautuvainen ja sietää poikkeuksellisia tai puuttuvia arvoja. Menetelmä kykenee myös havaitsemaan sellaisia hierarkkisia suhteita muuttujien arvojen välillä, joita perinteiset regressiomenetelmät eivät havaitse – tämä näkyy esimerkiksi menetelmän tuottamissa vastekäyrissä, joissa muuttujien kynnysarvot erottuvat selvästi. Menetelmän suurimmat ongelmat liittyvät juuri sen aineistolähtöisyyteen ja joustavuuteen, mikä voi johtaa esimerkiksi ylisovittamiseen huolimatta menetelmän sisäisistä keinoista välttää ylisovittamista (Merow ym. 2014). Käyttäjän tulee itse huolehtia esimerkiksi asetusten säätämisestä ja muuttujavalinnoista, mikä voi vaatia suurempaa työpanosta kuin mallien sovittaminen menetelmälähtöisten algoritmien avulla. Mallinnusprosessiin liittyy myös tietynlaista stokastisuutta mallien välillä johtuen ylisovittamisen kontrollointiin käytetyistä menetelmistä, joissa hyödynnetään ristiinvalidointia (Elith ym. 2008; Guisan ym. 2017: 210–212).

Mallien sovittamista kontrolloitiin säätelemällä mallin oppimisastetta (*learning rate*), päätöspuiden monimutkaisuutta ja määrää. Oppimisaste pidettiin matalana, jotta

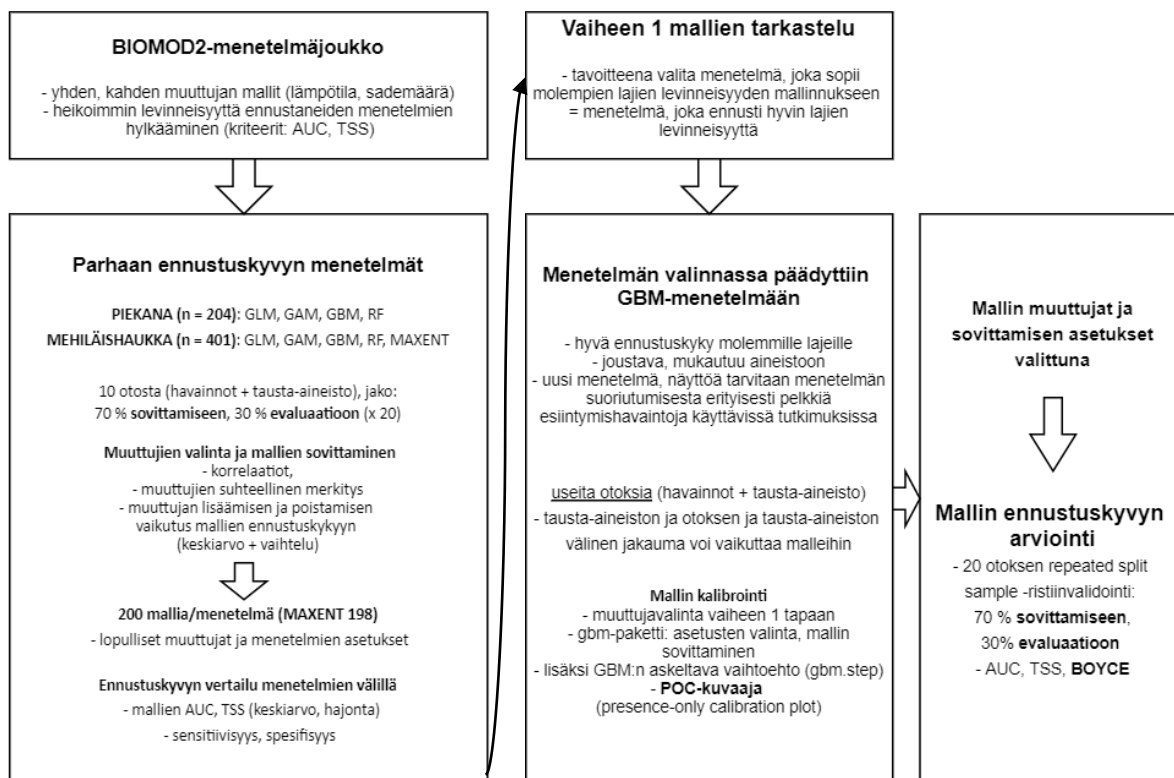
riittävän suuri määrä puita saatiin sovitettua malliin mallin parantamiseksi ja tasapainottamiseksi. Oppimisaste oli molempien lajien mallissa 0,001. Elith ynnä muut (2008) suosittelevat pitämään puiden monimutkaisuuden (*tree complexity / interaction depth*) lukujen kaksi ja kymmenen välillä liiallisen monimutkaisuuden rajoittamiseksi. Molemmilla lajeilla monimutkaisuus rajoitettiin lukemaan viisi, koska lukeman nostaminen ei näyttänyt tuovan malleille selvää lisäarvoa. Sovittamista suoritettiin **useilla eri havaintojen ja tausta-aineiston otoksilla**, jotta sopivat muuttujayhdistelmät ja asetukset voitiin määrittellä siten, että lopulliseksi malliksi ei valikoituisi malli, joka on riippuvainen esimerkiksi havaintojen ja tausta-aineiston poikkeuksellisesta maantieteellisestä jakaumasta. Muuttujien valinnan apuna käytettiin *dismon* askeltavaa *gbm.step*-funktioita. Muuttujayhdistelmille laskettiin otoskohtaiset VIF-arvot muuttujien välillä mahdollisesti olevien piilevien suhteiden havaitsemiseksi. Mallin sovittamisen arvioinnissa käytettiin myös visuaalista menetelmää, POC-kuvaajaa (*presence-only calibration plot*), joka ottaa huomioon ainoastaan esiintymishavainnot ja sopii näin erinomaisesti pelkkiä esiintymishavaintoja sisältävän aineiston mallinnusten arviointiin (Phillips & Elith 2010). GBM-menetelmällä levinneisyysmallin sovittaminen voi olla haastavaa ja POC-kuvaajan käyttö tukee sopivan mallin muodostamista. Mallien residuaaliarvot tarkistettiin spatiaalisen autokorrelaation varalta Moranin I:n avulla kymmenessä etäisyysluokassa (*ape*-paketti, Paradis ym. 2012).

Evaluaatiossa käytettiin TSS:n ja AUC:n lisäksi myös kolmatta arviointikriteeriä, Boycen indeksiä, joka arvioi ennusteiden laatua ainoastaan esiintymishavaintojen pohjalta soveltuen näin hyvin kyseisellä aineistolla sovitettujen mallien arviointiin (Boyce ym. 2002; Phillips & Elith 2010). Se jakaa ennusteet tasaisiin luokkiin ja ikään kuin vertaa havaintojen ja ennusteiden jakaumaa luokissa tilanteeseen, jossa kaikki havainnot on jaettu satunnaisesti näihin samoihin luokkiin (Hirzel ym. 2006). Muodostuvalle arvolle F_i (*predicted-to-expected ratio*) ja luokkien mediaanille lasketaan sen jälkeen Spearmanin järjestyskorrelaatiokerroin: mitä lähempänä tämä on lukua 1, sitä paremmin ennusteet vastaavat todellisten esiintymishavaintojen jakautumista evaluaatioaineistossa (Hirzel ym. 2006). Evaluaatio suoritettiin 20 otoksen *repeated split sample* -ristiininvalidointimenetelmällä (James ym. 2013), jossa 70 % aineistosta käytettiin mallin sovittamiseen ja 30 % mallin testaamiseen. Tausta-aineiston käyttö (Phillips ym. 2009) ja GBM-menetelmään liittyvä stokastisuus (Elith ym. 2008) hankaloittavat mallien luotettavuuden ja todellisen ennustuskyvyn arviointia.

Esiteltujen tutkimusvaiheiden (kuva 5) tarkoitus on selvittää, minkälaisia levinneisyysmallinnuksen menetelmiä on mahdollista hyödyntää, kun tavoitteena on ennustaa päiväpetolintujen levinneisyyttä Suomessa hyödyntäen pelkistä yksilöhavainnoista koostuvaa, matalaresoluutioista aineistoa, joka on kerätty verrattain suurelta tutkimusalueelta. Piekana ja mehiläishaukka toimivat esimerkkeinä edustaessaan levinneisyysalueen maantieteelliseltä levinneisyydeltään, habitaatinvalinnaltaan ja esimerkiksi ravintotottumuksiltaan toisistaan poikkeavia päiväpetolintulajeja. Samanaikaisesti huomiota kiinnitetään ennusteiden luotettavuuteen ja pohditaan, soveltuvatko havaintoaineisto ja käytetyt ympäristömuuttujat ylipäätään kyseisten lajien levinneisyysmallinnukseen Suomessa. Pääpaino tulosten raportoinnissa ja arvioinnissa on mallien erottelukyvyyssä aineistosta satunnaisotannalla valikointuneisiin erillisiin evaluaatio-otoksiin. Vaikka käytettävissä ei ole täysin mallinnuksesta erillistä, itsenäistä aineistoa mallien evaluaatioon, menetelmän, aineistojen ja esimerkiksi muuttujien valinnan merkitystä malleille voidaan tarkastella erikseen ja arvioida näin ennusteiden luotettavuutta käytettyyn aineistoon esimerkiksi paljastamalla eroja mallin ennustuskyvyyssä, kun havainto- ja tausta-aineistot muuttuvat.

Vaihe 1: BIOMOD2-mallinnus (Thuiller ym. 2009)

Vaihe 2: GBM-mallit (Elith ym. 2008)



Kuva 5. Tiivistelmä levinneisyysmallinnuksen vaiheista ja etenemisestä tutkielmassa. GBM = Generalized Boosted Model, RF = Random Forest, GLM = Generalized Linear Model, GAM = Generalized Additive Model, MAXENT = Maximum Entropy Modeling, AUC = Area Under the Curve, TSS = True Skill Statistic, BOYCE = Boycen indeksi.

Tulokset

Menetelmien ennustuskyvyn tarkastelu ja vertailu

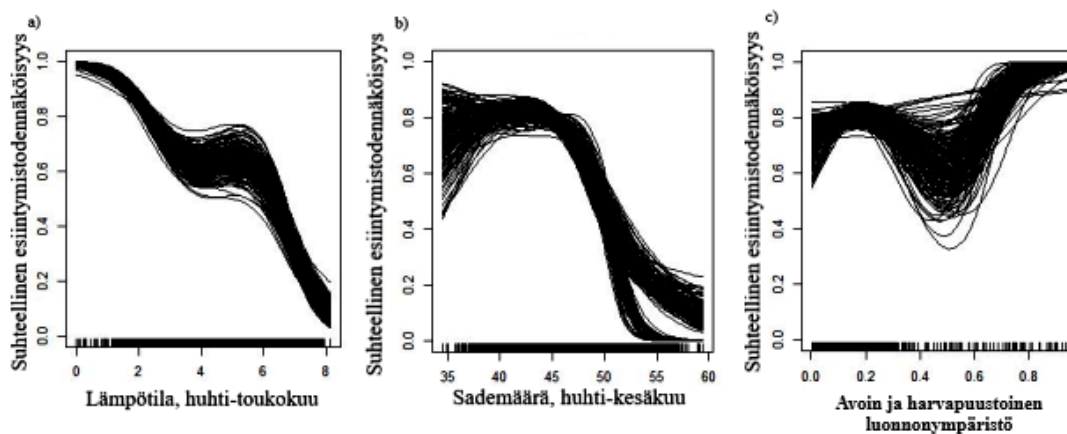
Menetelmien vertailussa piekanan levinneisyyttä ennustaviksi muuttujiksi valikoituivat huhti-toukokuun lämpötila, huhti-kesäkuun sademäärä ja avoimen ja harvapuustoisien luonnonympäristön määrä. Kevään keskilämpötilalla oli kaikkien menetelmien mukaan selvästi voimakkain merkitys piekanan levinneisyydelle (taulukko 4). Vaihtoehtoisten ajankohtien lämpötilojen käytölle ei ollut perusteita sillä niiden avulla luodut mallit ennustivat piekanan levinneisyyttä heikommin. Sademäärä oli seuraavaksi tärkein muuttuja, kun taas avoimen luonnonympäristön määrällä oli selvästi pienin merkitys. GAM-malleissa sateisuuden merkitys oli lähes lämpötilan veroinen. Huolimatta avoimen luonnonympäristön näennäisen heikosta merkityksestä piekanan esiintymiselle muuttujan lisääminen paransi mallien erottelukykyä verrattuna pelkkiin lämpötilan ja sateisuuden keskiarvoihin. Muiden maanpeitettä kuvaavien muuttujien merkitys oli heikko ja ne jätettiin pois menetelmien vertailusta.

Mallien vastekäyristä (kuva 6, s. 44, liite 2) voi tulkita, että piekanan suhteellinen esiintymistodennäköisyys oli korkeimmillaan matalissa lämpötiloissa ja laski lämpötilan noustessa. Myös korkeilla sademäärillä oli negatiivinen vaikutus piekanan esiintymiseen, mutta haitallinen vaikutus alkoi vasta sadannan ylitettyä tietyn raja-arvon. Avoimen ympäristön määrän kasvu näyttäisi mallien perusteella nostavan lajin suhteellista esiintymistodennäköisyyttä varsinkin suurimmilla muuttujan arvoilla. GBM- ja RF -mallit ennustivat muuttujille samansuuntaisia vasteita kuin GAM-mallit (liite 2). Muuttujien VIF-arvojen tarkastelu ei paljastanut multikollinearisuutta arvojen ollessa lähellä lukua 1. Suurin VIF-arvo oli lämpötilan 1,38.

Taulukko 4. Muuttujien suhteellinen merkitys piekanan levinneisyydessä menetelmittain. GBM = Generalized Boosted Model, RF = Random Forest, GAM = Generalized Additive Model.

Muuttuja	GBM	RF	GAM
Lämpötila, huhti-toukokuu	0,723	0,582	0,519
Sademäärä, huhti-kesäkuu	0,226	0,217	0,473
Avoim tai harvapuustoinen luonnonympäristö	0,066	0,136	0,032

BIOMOD2:lla vertailluista levinneisyysmallinnuksen menetelmistä GAM, GBM ja RF erottuivat selvästi muista erottelukyvyltään. Kyseisten menetelmien luomien mallien AUC- ja TSS-arvojen keskiarvojen erot ja vaihtelu **menetelmien välillä** olivat pieniä, vain joitain kymmenyksiä (taulukko 5, kuva 7, s. 45). Sensitiivisyyden ja spesifisyyden arvot olivat lähellä toisiaan ja vain RF:n kohdalla spesifisyys oli sensitiivisyyttä korkeampi (taulukko 6). Arvojen vaihteluväleistä, keskihajonnoista ja varsinkin laatikko-jana -kuvioista voidaan nähdä, että erot **yksittäisen menetelmän** mallien välillä olivat selviä mutta eivät keskimäärin hälyttävän suuria.



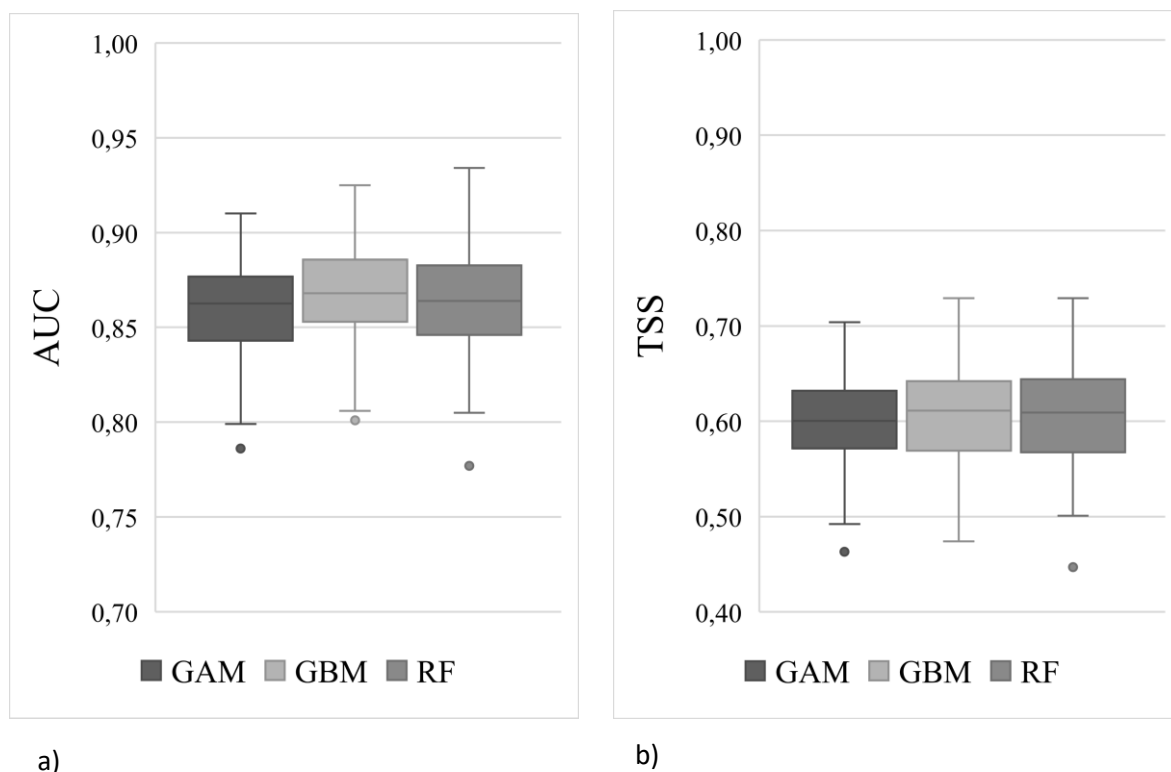
Kuva 6. GAM-vastekäyrät, mehiläishaukka a) touko-kesäkuun lämpötila, b) touko-kesäkuun sateisuus ja c) avoimen ja harvapuustoisien luonnonympäristön määrä.

Taulukko 5. Piekanan mallien (n = 200) erottelukyvyn arvojen tunnusluvut. AUC = Area Under the Curve, TSS = True Skill Statistic, GBM = Generalized Boosted Model, RF = Random Forest, GAM = Generalized Additive Model.

Arvo, menetelmä	Keskiarvo	Keskihajonta	Minimiarvo	Maksimiarvo
AUC, GBM	0,868	0,024	0,801	0,925
AUC, RF	0,864	0,025	0,777	0,934
AUC, GAM	0,860	0,022	0,786	0,910
TSS, GBM	0,609	0,052	0,474	0,729
TSS, RF	0,607	0,049	0,447	0,729
TSS, GAM	0,601	0,047	0,463	0,704

Taulukko 6. Piekana, sensitiivisyyden ja spesifisyyden keskiarvot malleissa (maxTSS, n = 200). Sensitiivisyys mittaa oikein ennustettujen esiintymisten osuutta ja spesifisyys oikein ennustettujen poissaolojen osuutta malleissa. GBM = Generalized Boosted Model, RF = Random Forest, GAM = Generalized Additive Model.

Arvo	GBM	RF	GAM
Sensitiivisyys	0,816	0,799	0,819
Spesifisyys	0,792	0,807	0,781



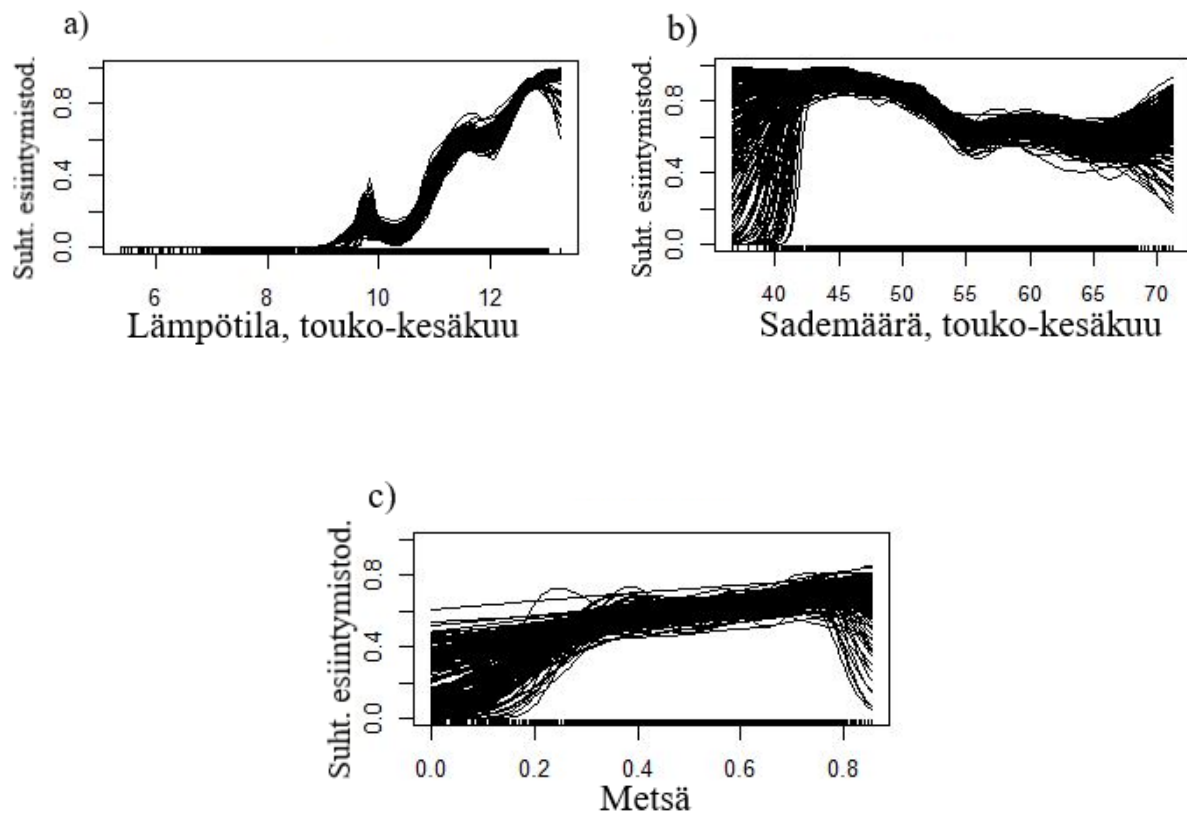
Kuva 7. Piekana, laatikko-jana -kuviot erottelukyvyn arvojen jakaumalle a) AUC (Area Under the Curve) ja b) TSS (True Skill Statistic). GBM = Generalized Boosted Model, RF = Random Forest, GAM = Generalized Additive Model. **Huomaa toisistaan eroavat y-akselin arvoasteikot.**

Mehiläishaukan kohdalla muuttujiksi valikoituivat touko-kesäkuun lämpötila, touko-kesäkuun sademäärä ja metsien määrä. Mallien mukaan lämpötilan merkitys mehiläishaukan levinneisyydelle oli huomattavasti suurempi kuin sadannalla tai metsien määrällä, joiden keskinäinen merkitys vaihteli menetelmittäin (taulukko 7, s. 46). Huolimatta sademäärän ja metsien määrän pienestä merkityksestä verrattuna lämpötilan keskiarvoon molempien sisällyttäminen malleihin paransi mallien keskimääräistä erottelukykyä.

GAM-vastekäyrien (kuva 8, s. 46) mukaan mehiläishaukan touko-kesäkuun lämpötilan kasvu nosti mehiläishaukan suhteellista esiintymistodennäköisyyttä: mehiläishaukka näyttää vaativan tietynsuuruisen lämpötilan loppukeväästä ja alkukesästä, jotta esiintyminen on ylipäättään mahdollista. Suurilla sademäärillä näytti olevan negatiivinen vaikutus myös mehiläishaukan esiintymiseen, mutta vaikutuksen voimakkuus ja suunta eivät ole yhtä selkeät kuin piekanalla. Metsien määrällä taas oli positiivinen vaikutus mehiläishaukan esiintymiseen. Myös muiden menetelmien tuottamat vastekäyrät olivat pääosin samansuuntaisia (liite 3). Muuttujissa ei havaittu multikollineaarisuutta suurimman VIF-arvon ollessa sadannan 1,43.

Taulukko 7. Muuttujien suhteellinen merkitys mehiläishaukan levinneisyydessä menetelmittäin (MAXENT-tulosta ei saatavilla kahdesta epäonnistuneesta mallista johtuen). GBM = Generalized Boosted Model, RF = Random Forest, GLM = Generalized Linear Model, GAM = Generalized Additive Model, MAXENT = Maximum Entropy Modeling.

Muuttuja	GBM	RF	GLM	GAM	MAXENT
Lämpötila, touko-kesäkuu	0,9854	0,8906	0,9845	0,9874	NA
Sademäärä, touko-kesäkuu	0,0548	0,1519	0,049	0,0639	NA
Metsä	0,038	0,1162	0,0628	0,0471	NA



Kuva 8. GAM-vastekäyrät, mehiläishaukka a) touko-kesäkuun lämpötila, b) touko-kesäkuun sateisuus ja c) metsien määrä.

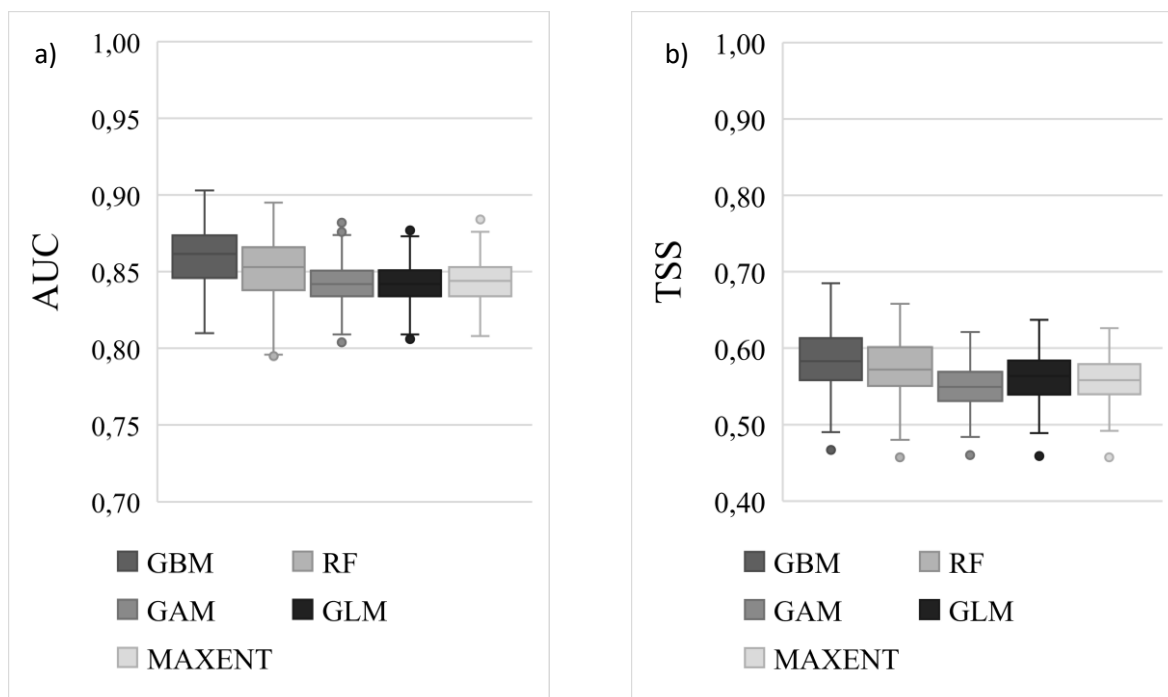
Mehiläishaukan kohdalla useat menetelmät tuottivat AUC- ja TSS-arvoiltaan vastaavia ennusteita käytössä olleella aineistolla. Mallien erottelukyky oli keskimäärin piekanan malleja heikompi mutta mallien ennusteiden välinen menetelmäkohtainen vaihtelu erottelukyvissä oli vähäisempää varsinkin TSS:n kohdalla (taulukko 8, kuva 9, s. 48). GBM ja RF erottuivat hieman paremmalla erottelukyvyllään mutta ero muihin menetelmiin ei ole suuri. Huomiota herättävää on sensitiivisyyden suuruus ja sen suhde spesifisyyteen mehiläishaukamalleissa (taulukko 9). Toisin kuin piekanalla mehiläishaukan mallien sensitiivisyys oli spesifisyyttä huomattavasti korkeampi ja sensitiivisyys sai suuremmat keskiarvot kuin piekanan vastaavat.

Taulukko 8. Mehiläishaukan mallien erottelukyvyn arvojen tunnusluvut (n = 200, MAXENT-mallit: n = 198). AUC = Area Under Curve, TSS = True Skill Statistic, GBM = Generalized Boosted Model, RF = Random Forest, GLM = Generalized Linear Model, GAM = Generalized Additive Model, MAXENT = Maximum Entropy Modeling.

Arvo, menetelmä	Keskiarvo	Keskihajonta	Minimiarvo	Maksimiarvo
AUC, GBM	0,860	0,018	0,810	0,903
AUC, RF	0,851	0,020	0,795	0,895
AUC, GAM	0,843	0,015	0,804	0,882
AUC, GLM	0,843	0,014	0,806	0,882
AUC, MAXENT	0,844	0,015	0,808	0,884
TSS, GBM	0,585	0,038	0,467	0,685
TSS, RF	0,576	0,038	0,457	0,658
TSS, GAM	0,550	0,029	0,460	0,621
TSS, GLM	0,563	0,030	0,459	0,637
TSS, MAXENT	0,558	0,031	0,457	0,626

Taulukko 9. Mehiläishaukka, sensitiivisyyden ja spesifisyyden keskiarvot malleissa (maxTSS, n = 200, MAXENT-mallit: n = 198). Sensitiivisyys mittaa oikein ennustettujen esiintymisten osuutta ja spesifisyys oikein ennustettujen poissaolojen osuutta malleissa. GBM = Generalized Boosted Model, RF = Random Forest, GLM = Generalized Linear Model, GAM = Generalized Additive Model, MAXENT = Maximum Entropy Modeling.

Arvo	GBM	RF	GLM	GAM	MAXENT
Sensitiivisyys	0,886	0,871	0,904	0,908	0,893
Spesifisyys	0,697	0,703	0,658	0,641	0,664



Kuva 9. Mehiläishaukka, laatikko-jana -kuviot erottelukyvyn arvojen jakaumalle a) AUC (Area Under Curve) ja b) TSS (True Skill Statistic). GBM = Generalized Boosted Model, RF = Random Forest, GLM = Generalized Linear Model, GAM = Generalized Additive Model, MAXENT = Maximum Entropy Modeling. Huomaa toisistaan eroavat y-akselin arvoasteikot.

Mallien sovittaminen *Generalized Boosted Model* -menetelmällä ja luotettavuuden arviointi

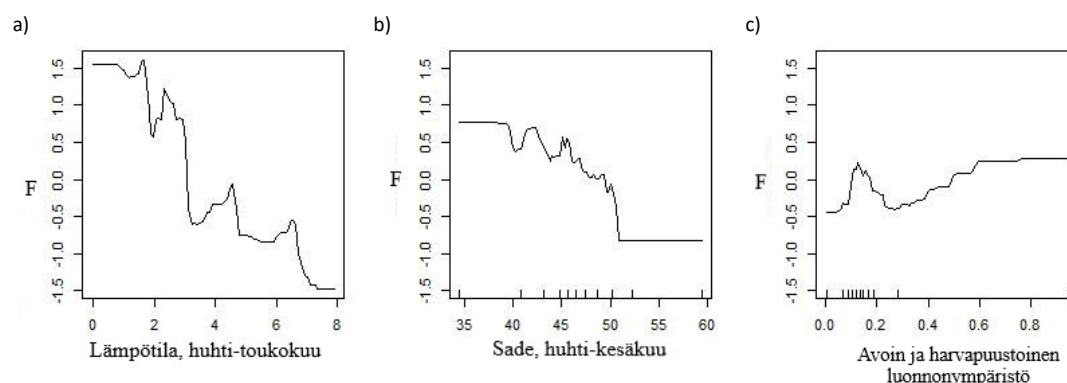
Piekanan GBM-malliin valikoituivat muuttujiksi samat muuttujat kuin ensimmäisessä vaiheessa: huhti-toukokuun lämpötila, huhti-kesäkuun sademäärä ja avoimen harvapuustoisien luonnonympäristön määrä. Mehiläishaukan malliin lisättiin ensimmäiseen vaiheeseen verrattuna neljäs muuttuja, maatalousalueiden määrä. Muuttujat olivat näin touko-kesäkuun lämpötila, touko-kesäkuun sademäärä, metsien ja maatalousalueiden määrä. Lämpötilat olivat selvästi merkityksellisimmät muuttujat molemmille lajeille (taulukko 10). Piekanalle sademäärä oli hieman avoimen luonnonympäristön määrää tärkeämpi. Mehiläishaukalle sademäärä, maatalouden ja metsien määrä olivat lähes yhtä merkityksellisiä.

Lämpötilan ja sateen määrä vaikuttivat negatiivisesti piekanan esiintymiseen (kuva 10). Kun avoimen ja harvapuustoisien luonnonympäristön määrä oli suuri, piekana suhteellinen esiintymistodennäköisyys vaikutti olevan suurimmillaan. Lämpötilan merkitys laski verrattuna ensimmäiseen vaiheeseen ja avoimen luonnonympäristön määrän merkitys puolestaan nousi. Lämpötilan määrä vaikutti voimakkaasti myös

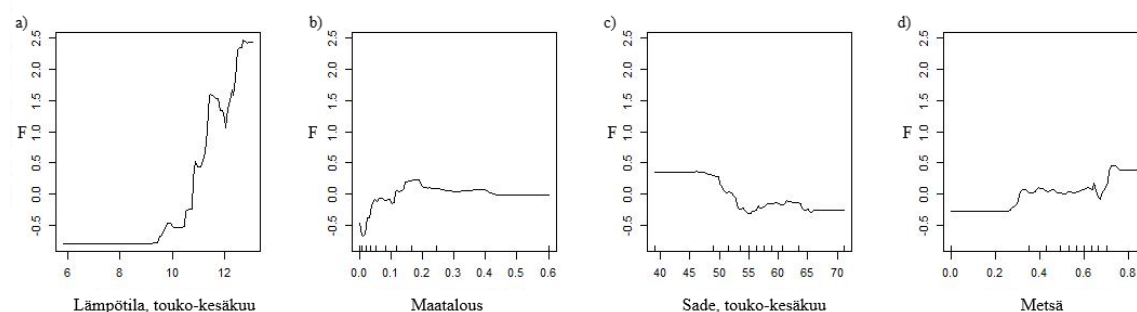
mehiläishaukkaan, mutta päinvastoin kuin piekanalla, korkeat lämpötilat olivat suotuisimpia sen esiintymiselle (kuva 11). Sademäärä vaikutti piekanan tapaan negatiivisesti mutteivät yhtä voimakkaasti kuin piekana. Maatalousalueiden ja metsien määrällä näytti olevan positiivinen vaikutus mehiläishaukan levinneisyydelle. Muuttujien välillä ei havaittu multikollineaarisuutta (VIF, taulukko 10).

Taulukko 10. Muuttujien suhteellinen merkitys mallille prosentteina ja VIF-arvo (Variance Inflation Factor).

	Muuttuja	Suhteellinen merkitys (%)	VIF
Piekana	Lämpötila, huhti-toukokuu	57,8	1,56
	Sademäärä, huhti-kesäkuu	24,1	1,28
	Avoin tai harvapuustoinen luonnonympäristö	18,1	1,38
Mehiläishaukka	Lämpötila, touko-kesäkuu	62,0	1,79
	Maatalous	15,4	1,84
	Sademäärä, touko-kesäkuu	12,2	1,44
	Metsä	10,4	1,35



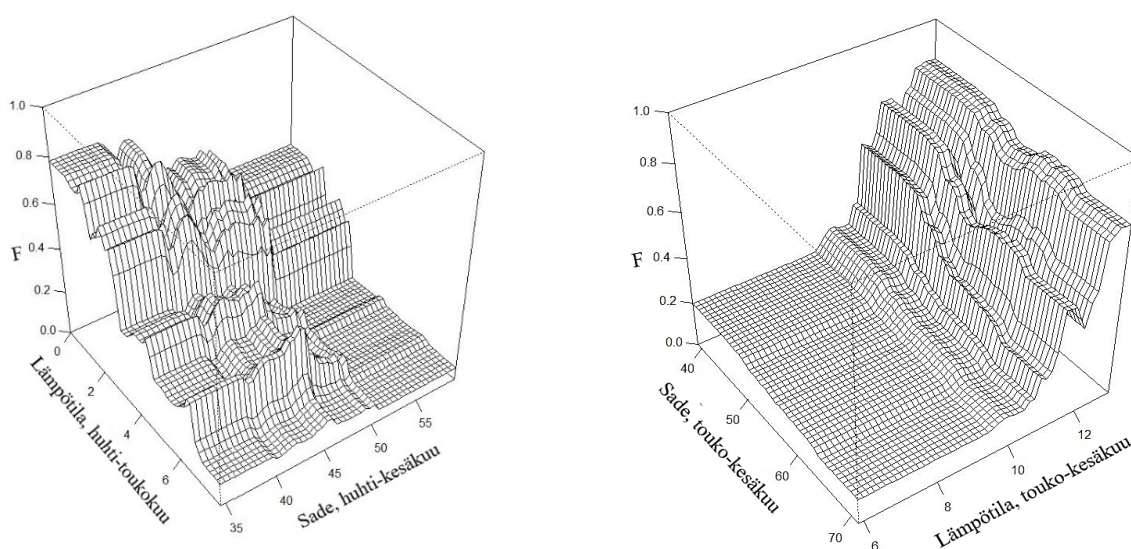
Kuva 10. Vastekäyrät (*partial dependence plots*), **piekana**, a) huhti-toukokuun lämpötila (°C), b) huhti-kesäkuun sademäärä (mm) ja c) avoimen ja harvapuustoisin luonnonympäristön määrä (%). Y-akseli logaritminen.



Kuva 11. Vastekäyrät (*partial dependence plots*), **mehiläishaukka**, a) touko-kesäkuun lämpötila (°C), b) maatalousalueiden määrä (%), c) touko-kesäkuun sademäärä (mm) ja d) metsien määrä (%). Y-akseli logaritminen.

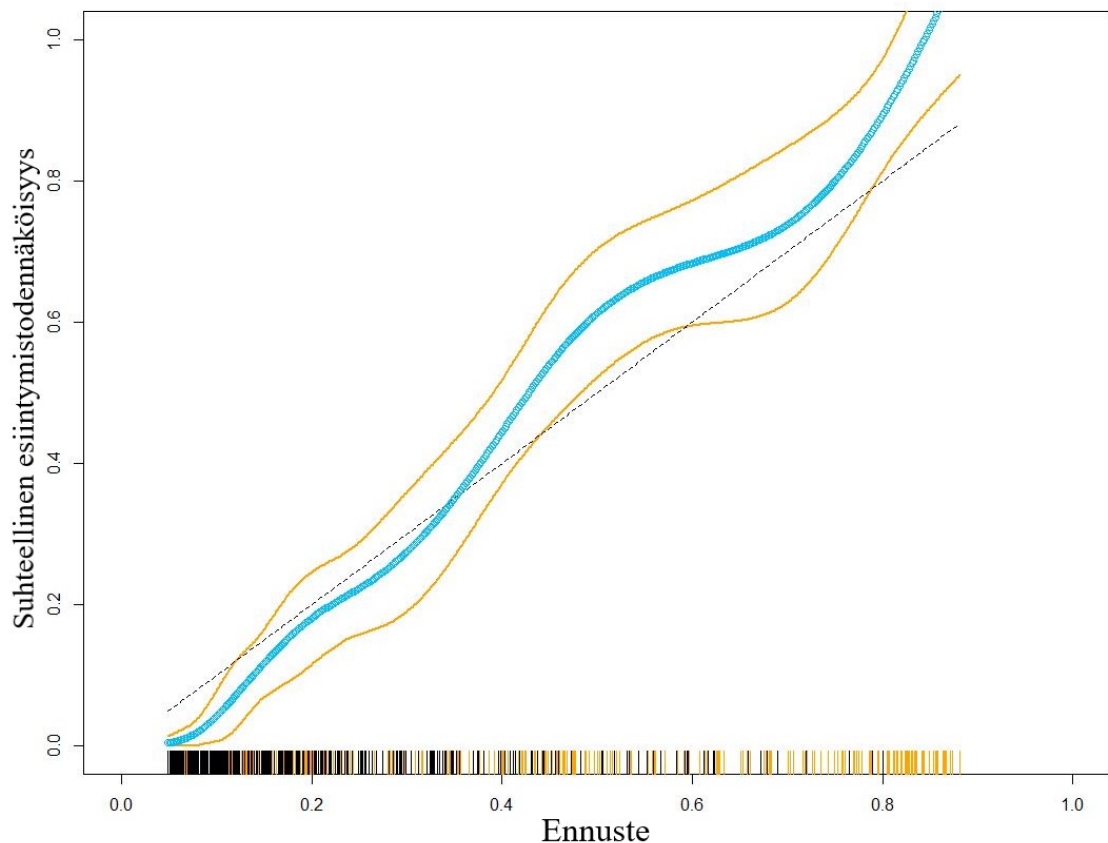
GBM-mallien mukaan lämpötilan ja sademäärän yhteisvaikutus vaikutti voimakkaasti piekanan ja mehiläishaukan suhteelliseen esiintymistodennäköisyyteen. Lämpötilan merkitys molempien lajien levinneisyyteen erottuu selvästi lämpötilan ja sademäärän kolmiulotteisista kuvaajista (kuva 12). Kun lämpötila pysytteli riittävän matalalla, piekanan esiintymistodennäköisyys ei laskenut alle 0,5:n sademääränkään kasvaessa. Lämpötilan noustessa sademäärän lievä kasvu voi jopa nostaa piekanan esiintymistodennäköisyyttä, kunhan sademäärä pysyy kohtalaisena. Mehiläishaukan esiintyminen näytti vaativan noin kymmenen celsiusasteen keskilämpötilan touko-kesäkuussa, ja tuon raja-arvon ylityksen jälkeen esiintymistodennäköisyys kasvoi tasaisesti. Sademäärän kasvu rajoitti esiintymistä sitä vähemmän mitä korkeampia lämpötilat ovat. Kaiken kaikkiaan molempien lajien vasteet ympäristömuuttujiin noudattivat ensimmäisessä vaiheessa havaittua suuntaa ja voimakkuutta.

POC-kuvaajien perusteella molempien lajien mallit ennustivat lajin esiintymisiä tasaisesti koko aineistossa (kuva 13 s. 51, kuva 14, s. 52). Yli- ja alisovittamista oli havaittavissa ja luottamusvälit kasvoivat paikoitellen, mutta kuvaajien mukaan mallit suoriutuivat pääosin tasaisesti mallin sovittamisessa. Mallien selitysasteet olivat hyvät aineiston puutteet huomioon ottaen. Ne olivat molemmilla lajeilla myös samaa luokkaa (taulukko 11, s. 52). Huolimatta mehiläishaukan mallin pienemmistä TSS- ja AUC-arvoista sen Boycen indeksilukujen keskiarvo on piekanan vastaavaa korkeampi ja sen hajonta on pienempi.

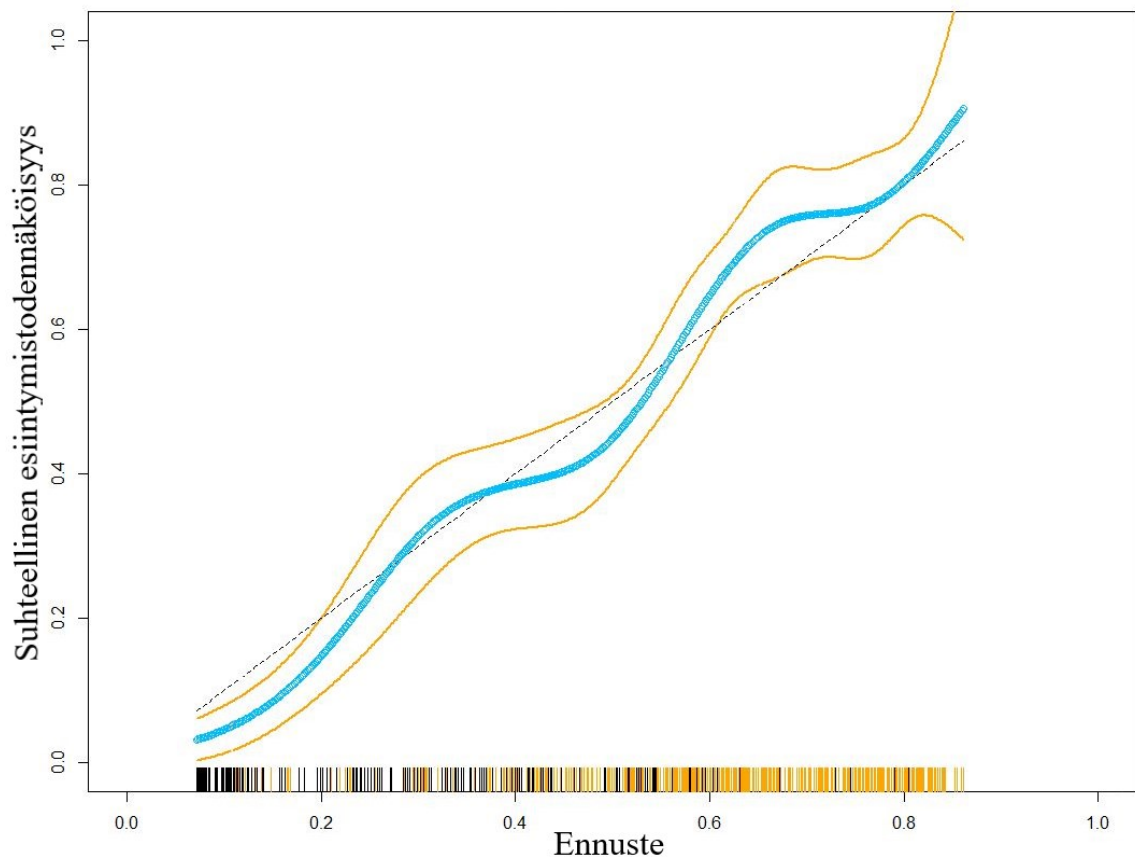


Kuva 12. Kolmiulotteinen kuvaesitys mallien kausittaisten lämpötilojen (°C) ja sademäärien (mm) vuorovai-
kutuksesta lajin suhteelliseen esiintymistodennäköisyyteen, a) **piekana** ja b) **mehiläishaukka**.

Mehiläishaukan malli näyttäisi siis ennustavan piekanan mallia paremmin lajin esiintymistä, kun ennustaminen tausta-aineiston maantieteelliseen jakaumaan jätetään huomiotta ja arvioidaan pelkästään mallin kykyä ennustaa todellisten esiintymishavaintojen jakautumista Suomessa. TSS- ja AUC -arvot vastaavat ensimmäisen vaiheen GBM-mallien arvoja molemmilla lajeilla, ollen keskiarvoltaan hieman ensimmäisen vaiheen keskiarvoja korkeampia. Arvoissa on jonkin verran hajontaa, mutta ei mitään sellaista, mikä ei olisi odotettavissa tällä aineistolla. Mallien kalibraation tarkastelu ja evaluaatio ei paljasta virhelähteitä, jotka vaikuttaisivat ennusteiden uskottavuuteen merkittävästi. Molempien lajien mallien residuaaleissa oli positiivista spatiaalista autokorrelaatiota Moranin I:n ollessa piekanalla 0,035 ja mehiläishaukalla 0,043. Spatiaalinen autokorrelaatio oli tilastollisesti merkitsevää (p -arvo $< 0,001$) ja sitä havaittiin molemmilla lajeilla neljässä ensimmäisessä etäisyysluokassa (liite 4).



Kuva 13. POC-kuvaaja (*presence-only calibration plot*), **piekana**. Sinisen kalibrointikäyrän tulisi seurata trendiviivaa. Kalibrointikäyrää ympäröivät oranssit viivat kuvaavat luottamusväliä (keskihajonta ± 2). Keltaiset pystyviivat näyttävät ennusteen arvot havaintoruuduissa, mustat ruuduissa, joissa lajia ei ole havaittu.



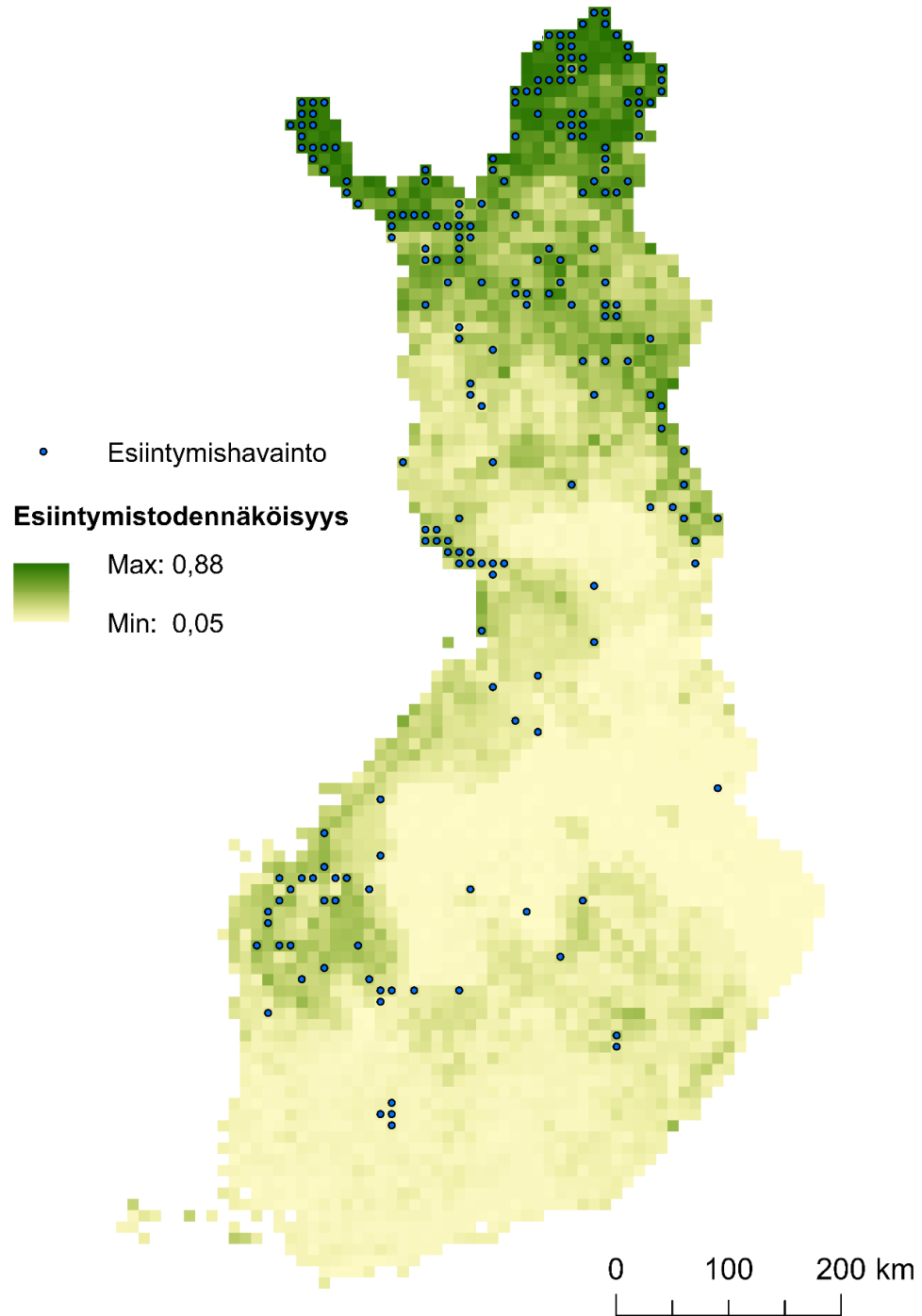
Kuva 14. POC-kuvaaja (*presence-only calibration plot*), **mehiläishaukka**. Sinisen kalibrointikäyrän tulisi seurata trendiviivaa. Kalibrointikäyrää ympäröivät oranssit viivat kuvaavat luottamusväliä (keskihajonta ± 2). Keltaiset pystyviivat näyttävät ennusteen arvot havaintoruuduissa, mustat ruuduissa, joissa lajia ei ole havaittu.

Taulukko 11. GBM-mallien (*Generalized Boosted Model*) tulokset ja evaluaatioarvot (keskiarvot, suluissa keskihajonta). CV korr. = ristiinvalidoitu korrelaatiokerroin, Boycen indeksi, TSS = True Skill Statistic, AUC = Area Under Curve.

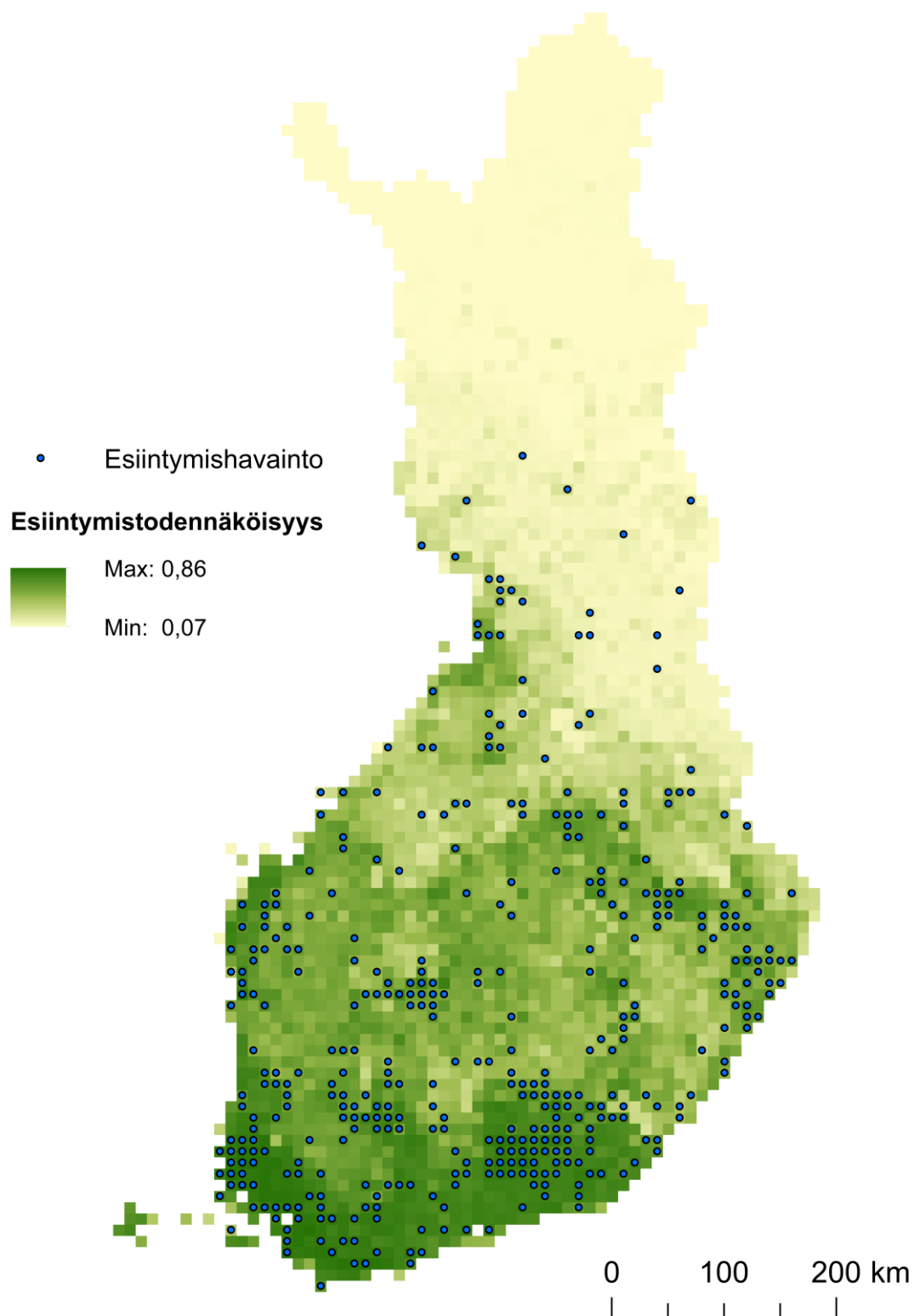
	Puiden määrä	Luokittelu- virhe	CV korr.	Boyce	TSS	AUC
Piekana	3350	0,907	0,544 (0,027)	0,911 (0,047)	0,668 (0,039)	0,904 (0,017)
Mehiläis- haukka	2850	1,068	0,519 (0,017)	0,94 (0,024)	0,621 (0,030)	0,883 (0,012)

Karttaesitykset (kuva 15, kuva 16, s. 54) kuvaavat GBM-mallien ennusteita piekanan ja mehiläishaukan suhteellisesta esiintymistodennäköisyydestä Suomessa. Mitä suurempina arvot näyttäytyvät kartalla sitä sopivampia kyseiset habitaatit ovat. Pienet arvot taas kertovat suhteessa heikkolaatuisemmista habitateista. Korkeimmat esiintymistodennäköisyydet painottuvat piekanalla pohjoisimpaan Lappiin, kun taas mehiläishaukan

esiintymistodennäköisyys näyttää laskevan siirryttäessä suuria arvoja sisältävästä eteläisimmästä Suomesta kohti pohjoista. Ennusteet sopivat hyvin esiintymishavaintojen havaittuun jakautumiseen.



Kuva 15. Ennusteet **piekanan** suhteellisesta esiintymistodennäköisyydestä Suomessa ja esiintymishavainnot, resoluutio 10 km x 10 km.



Kuva 16. Ennusteet **mehiläishaukan** suhteellisesta esiintymistodennäköisyydestä Suomessa ja esiintymishavainnot, resoluutio 10 km x 10 km.

Pohdinta

Tarkasteltujen menetelmien ennustuskyvyyssä ei havaittu suuria eroja

Menetelmien vertailu näytti, että piekanan ja mehiläishaukan levinneisyyttä voi mallintaa useilla eri menetelmillä eikä yksittäinen menetelmä suoriutunut selvästi muita paremmin. Regressiomenetelmät yhdessä satunnaismetsän (*random forest, RF*) ja GBM-menetelmän kanssa ovat hyvä lähtökohta päiväpetolintujen levinneisyyden mallinnukseen yksinkertaisella aineistolla ja matalalla resoluutiolla. MAXENT-algoritmi ei esimerkiksi erottunut muiden menetelmien joukosta, vaikka menetelmästä on tullut *presence only* -aineistojen parissa suosittu menetelmä sen jälkeen, kun Phillips ynnä muut toivat sen levinneisyysmallinnuksen pariin vuonna 2006 (Qiao ym. 2015). Tarkastelussa ei paljastunut merkittäviä eroja erilaisten menetelmien erottelukykyjen välillä. Vaikka toisen vaiheen lopullisissa malleissa käytettiin GBM-menetelmää, GLM ja GAM olisivat voineet tuottaa lähes samantaisia ennusteita lajien levinneisyydestä. Tämä vahvistaa esimerkiksi Qiaon ynnä muiden (2015) näkemystä siitä, että levinneisyysmallinnusta aloitettaessa ei tulisi sokeasti valita aiemmissa tutkimuksissa hyväksi havaittuja menetelmiä vaan tarkastella myös muita, vaihtoehtoisia menetelmiä ja pohtia ennen mallinnuksen aloitusta esimerkiksi, millainen aineisto on käytössä ja mikä mallinnuksen tavoite on. Aineistossa ei vaikuttanut olevan sellaista painottuneisuutta, joka olisi suosinut tietyn menetelmän käyttöä ylitse muiden. Päiväpetolintujenkin kohdalla tulisi pohtia tarkasti aineiston mahdollisuuksia ja rajoitteita tutkimuksen tavoitteen näkökulmasta.

Lajien AUC-arvot olivat Araújo ynnä muiden (2005) levinneisyysmallinnukseen esittämällä mittakaavalla keskimäärin ”hyvät” (AUC: 0,80 – 0,90). Myös TSS-arvot olivat hyvät (Allouche ym. 2006). Mallien erottelukykyä ei ole mielekäästä verrata suoraan aiempiin tutkimuksiin, joiden aineistot ja tutkimuksenasettelut poikkeavat tässä tutkielmassa käytetyistä. Virkkala ynnä muut (2005) tosin käyttivät vastaavaa *presence-only* -aineistoa, samaa resoluutiota ja vastaavia muuttujia mallintaessaan suolintujen levinneisyyttä GAM-menetelmällä. Kyseisessä tutkimuksessa erottelukyvyn arviointiin käytetyn AUC-arvot vaihtelivat kymmenellä lajilla välillä 0,702 – 0,947. AUC- ja TSS-arvojen lisäksi ensimmäisessä vaiheessa oli hyvä tarkastella spesifisyyden ja sensitiivisyyden arvoja, jotka vaikuttavat itse AUC:n ja TSS:n suuruuteen. Piekanan kohdalla molemmat arvot olivat lähes samat (~ 0,8) kaikilla menetelmillä mutta mehiläishaukalla sensitiivisyys oli huomattavasti spesifisyyttä korkeampi. Koska emme voi mitata mallien ennustuskkyä ruutuihin, joissa

lajia ei ollut tavattu (tiedämme vain, missä ruuduissa yksilöitä on havaittu), TSS ja AUC voivat johtaa meitä harhaan. Mallien sensitiivisyyden arvo sen sijaan on luotettavampi mittari, koska se mittaa mallin ennustuskykyä nimenomaan yksilöhavaintoihin nähden. Korkea sensitiivisyys viittaisi siihen, että mallit saattoivat ennustaa mehiläishaukan levinneisyyttä todellisuudessa jopa paremmin kuin piekanan mallit, vaikka TSS-arvo jäikin piekanan malleja pienemmäksi. Tätä havaintoa tukee myös mehiläishaukan korkeampi Boycen indeksi GBM-menetelmillä vaiheessa kaksi.

GBM-mallien AUC- ja TSS-arvot olivat toisessa vaiheessa hieman ensimmäisen vaiheen keskiarvoa korkeampia. Boycen indeksin keskiarvo oli molempien lajien malleilla lähellä maksimiarvoa 1, joten mallien erottelukyky oli todella hyvä. POC-kuvaajat tukevat mallien käyttöä huolimatta paikoittaisesta ylisovittamisesta. Boycen indeksin ja POC-kuvaajien tarkastelu ja AUC- ja TSS-arvojen samansuuntaisuus sekä ensimmäisessä että toisessa vaiheessa viittaa siihen, että AUC- ja TSS-arvoja pystyi pitämään ainakin suuntaa antavina piekana- ja mehiläishaukka-aineistoilla huolimatta niiden epäsopevuudesta aineistoon nähden. Mallien ennustuskyky oli siis hyvä. Aineiston resoluutio (McPherson ym. 2006), havaintojen luonne ja luotettavuus (Hijmans 2012), muuttujat (Austin 2002; Austin & Van Niel 2011) ja tausta-aineiston jakauma (Phillips ym. 2009) ynnä muut aineistoon liittyvät erityispiirteet voivat vaikeuttaa mallinnusta johtaen esimerkiksi yli- tai alisovittamiseen ja siten ennustuskyvyn vaihteluun (Elith & Leathwick 2009; Guisan ym. 2017). Yksittäisen tekijän todellista vaikutusta vaihtelussa voi olla vaikea tai lähes mahdotonta erottaa. Mehiläishaukan piekanaa suurempi otoskoko esimerkiksi on tekijä, joka epäilemättä vaikutti positiivisesti lajin mallien luotettavuuteen, vaikka todellista vaikutusta on vaikea mitata ilman lisäanalyysejä (Thibaud ym. 2014).

Ympäristömuuttujien tarkastelu voi paljastaa heikot levinneisyysmallit

Muuttujien tarkastelu on mielekäs lähtökohta arvioitaessa mallien luotettavuutta (Guisan & Thuiller 2005; Austin & Van Niel 2011). Piekanan levinneisyysmalleihin valikoituneet muuttujat ja niiden vaikutuksen suunta olivat lajin ekologian ja aineiston esiintymishavaintojen sijainnin näkökulmasta uskottavat. Piekanan levinneisyys painottuu pohjoiseen Suomeen, jossa lämpötilat ja sateisuus ovat vähäisempiä kuin etelässä niin vuositasolla kuin myös lajin kevätmuuton ja pesinnän aloituksen aikaan huhti-toukokuussa (Kersalo & Pirinen 2009). Pesimähavaintoja on tehty runsaasti myös Pohjanmaan rannikolla, jolla sadanta on keväällä ja alkukesästä esimerkiksi Itä-Suomea selvästi vähäisempää (liite 5). Mallien

ennusteissa sekä lämpötilan että sademäärän kasvu laskivat piekanan suhteellista esiintymistodennäköisyyttä, mutta avoimen ympäristön määrän vaikutus oli positiivinen.

Lämpötilan suhteellinen merkitys oli selvästi voimakkain piekanan levinneisyyttä selittävistä muuttujista. Fysiologisesta näkökulmasta tarkasteltuna ei tietenkään ole odotettavaa, että keväisten keskilämpötilojen nousu vaikuttaisi negatiivisesti lintulajin esiintymiseen (Root 1988; Austin 2002; Barbet-Massin ym. 2012). Lajin esiintymiset (kuva 15) kuitenkin painottuvat matalien lämpötilojen Lappiin sekä Pohjanmaan rannikkoseuduille, jolla Pohjanlahden vaikutus (Kersalo & Pirinen 2009: 14) pitää lämpötilat verrattain matalina pitkälle kesään saakka. Jos kerran matalat lämpötilat eivät vaikuta negatiivisesti piekanan pesintään ja sitä kautta lajin levinneisyyteen, niin miksi lämpötilojen nousu laskee lajin esiintymistodennäköisyyttä? Todennäköinen syy on lämpötilan vaikutuksessa habitaattien muodostumiseen ja niihin lajeihin, jotka vaikuttavat eri tavoin piekanan esiintymiseen (Root 1988; Laaksonen & Lehikoinen 2013; Terraube ym. 2016).

Lämpötila vaikuttaa yhdessä sateisuuden kanssa kasvillisuuteen, esimerkiksi metsien ja niiden puulajien biomassan määrään ja maantieteellisen jakautumiseen sekä metsänrajan muodostumiseen (Prentice ym. 1992; Myneni ym. 1997). Metsien peittävyys ja biomassan määrä pienenevät lämpötilan laskiessa ja esimerkiksi pohjoisimmassa Lapissa esiintyy metsien sijaan pelkkää tunturikoivikkoa ja kasvillisuus koostuu varvuista ja pensasmaisista kasveista (Virkkala ym. 2000). Lapissa myös soiden osuus luonnonympäristöistä on suuri johtuen eteläistä Suomea vähäisemmästä haihdunnasta ja ihmisvaikutuksesta (Virkkala ym. 2000). Piekana suosii harvapuustoisia habitaatteja (Gensbøl 1995); lämpötilan vaikutus erilaisten habitaattien muodostumiseen voi siis osaltaan selittää havaittua keskilämpötilojen vaikutusta piekanan esiintymistodennäköisyyteen.

Saaliseläinten esiintymisen ja määrän epäillään vaikuttavan piekanan levinneisyyteen ja populaatiokokoon (Reif ym. 2001; Terraube ym. 2015). Talven sääolosuhteet – erityisesti lämpötila – vaikuttavat voimakkaasti piekanan suosiman ravinnon, myyrien populaatioihin esimerkiksi lumipeitteen määrän (Terraube ym. 2015) ja myyrien ravintona käytämien kasvien selviytymisen (Korpela ym. 2013) kautta. Korpelan ynnä muiden mukaan talviaikaisen sään ja lumipeitteen merkitystä on kuitenkin saatettu liioitella ja kasvukauden aikaisilla lämpötiloillakin on merkitystä myyrien populaatiosykleille esimerkiksi ravinnon saatavuuden myötä. Suomessa lämpötilojen jakautuminen alueittain säilyy samankaltaisena eri vuodenaikoina siirryttäessä etelästä kohti pohjoista, myös keväällä ja alkukesästä (Kersalo & Pirinen 2009). Talvisen lumipeitteen määrät ja ajoittuminen puolestaan

noudattelevat pitkälti talven ja koko vuoden keskilämpötiloja (Kersalo & Pirinen 2009: 14–15). Mitä kylmempi alue on talvella sitä suurempi osa talvisesta sateesta tulee lumena. Piekana kykenee hyödyntämään huonoina myyrävuosina vaihtoehtoista ravintoa (Reif ym. 2001), mutta on todennäköistä, että myyrien saatavuus ohjaa vuosittain pesinnän onnistumista (Terraube ym. 2015, 2016) ja näin myös lajin levinneisyyttä.

Saaliseläinten ohella piekanaa vaikuttavia lajeja voivat olla muut samasta ravinnosta ja pesimäpaikoista taistelevat lajit, kuten muut petolinnut (Hakkarainen ym. 2004; Krüger ym. 2004; Ritchie & Johnson 2009). Erikoistuminen haastavimpiin ympäristöihin ei välttämättä minimoi kilpailun vaikutuksia: Potapov (1997) on esimerkiksi esittänyt, että piekanaa kookkaampi pohjoinen laji, tunturipöllö (*Bubo scandiacus*), rajoittaa piekanan mahdollisuutta hyödyntää täysimittaisesti myyräkantojen huippuvuosia Siperian tundralla saalistaessaan samaa ravintoa. Toisaalta lajin vaarana on myös joutuminen itse saaliiksi suuremman petolinnun toimesta (Ritchie & Johnson 2009). Esimerkiksi piekanan kanssa samaan sukuun kuuluvan hiirihaukan (*Buteo buteo*) on havaittu välttävän kanahaukkayksilöiden (*Accipiter gentilis*) habitaateilla pesimistä paitsi kilpailun vuoksi myös kanahaukan saaliiksi joutumisen uhasta (Björklund ym. 2016). Myös samoissa pesimäympäristöissä viihtyvät pöllölajit kilpailevat sopivista pesimäpaikoista ja myyräravinnosta päiväpetolintujen kanssa (Terraube ym. 2016). Huuhkajan (*Bubo bubo*) läsnäolo taas voi vaikuttaa sekä hiiri- että kanahaukan habitaatinvalintaan saalistusuhan vuoksi (Mueller 2016). Koska piekanan fysiologia ja elintottumukset muistuttavat lähisukulaistaan hiirihaukkaa (Sylvén 1978), on todennäköistä, että molempiin lajeihin vaikuttavat samat bioottiset suhteet lajien esiintyessä samassa habitaatissa. Piekanan sopeutuminen viileän ilmaston metsäntörmille alueille voikin olla seurausta sen kyvyttömyydestä kilpailla metsien petolintujen kanssa pesimäpaikoista ja ravinnosta (Carothers & Jaksic 1984; Pigot & Tobias 2013). Hiiri- ja kanahaukka pesivät piekanaa harvalukuisempina Lapissa, erityisesti maan pohjoisrajalla, jossa piekana pesi lukuisissa havaintoruuduissa (Valkama ym. 2011).

Sadanta vaikuttaa lämpötilan rinnalla esimerkiksi vesistöihin, kasvillisuuden määrään, laatuun ja sen maantieteelliseen jakautumiseen, maaperän ravinteiden määrään ja hajottajien toimintaan sekä soiden muodostumiseen (Myneni ym. 1997). Korkeiden huhtikesäkuun sademäärien negatiivinen vaikutus piekanan esiintymiseen voi olla seurausta sadannan vaikutuksesta habitaattien muodostumiseen ja jakautumiseen sekä sellaisista haitallisista vaikutuksista, joita runsaat sateet voivat aiheuttaa piekanan ja muiden päiväpetolintujen pesinnälle (Lehikoinen ym. 2009; Ancil ym. 2014; Terraube ym. 2015). Vähäiset

sateet ovat edesauttaneet piekanalle ominaisten metsättömien habitaattien muodostumista erityisesti Pohjois-Lapissa mutta myös muualla maassa (Virkkala ym. 2000).

Kevään ja alkukesän runsaat sateet ja viileä sää altistavat päiväpetolintujen poikaset hypotermialle (Kostrewa 1989; Krüger ym. 2004; Terraube ym. 2015). Sateisuus voi vaikeuttaa myös saalistusta saaliseläinten saatavuuden ja höyhenpiteen kastumisen myötä (Dawson & Bortolotti 2000; Robinson ym. 2017). Piekanan suosimat kalliot ovat erityisen suojattomia sateella (Mearns & Newton 1988). Sademäärällä on havaittu olevan negatiivinen vaikutus päiväpetolintujen pesintään myös Suomessa (Lehikoinen ym. 2009, 2013) ja puuttomalla tundralla, piekanan pohjoisimpia levinneisyysalueita muistuttavassa ympäristössä (Ancitil ym. 2014; Robinson ym. 2017). Pesinnällä on olennainen vaikutus lajien populaatiokokoon, pesimäpaikan valintaan ja näin lajin levinneisyyteen lyhyellä ja pitkällä aikavälillä (Pulliam 2000). Sadanta voi vaikuttaa myös myyrien saatavuuteen (Terraube ym. 2016).

Sademäärän vaikutus jäi lämpötilan vaikutusta pienemmäksi. Keskiarvoltaan viileimmillä alueilla runsaimmatkaan sateet eivät laskeneet piekanan suhteellista esiintymistodennäköisyyttä alle $\sim 0,5:n$ (kuva 12). Toisaalta niillä alueilla, joilla lämpötilan kasvu laski piekanan esiintymistodennäköisyyttä, vasta sadannan selvä kasvu laski esiintymistodennäköisyyden lähelle nollaa. Itä-Suomi on huhti-kesäkuussa selvästi sateisempi kuin pohjoinen Lappi ja Pohjanmaan rannikkoseutu (liite 5). GBM-mallin ennustekartassa (kuva 15) Itä-Suomi näyttäytyy piekanalle epäsuotuisana ympäristönä yhdessä lämpimän ja sateisen Lounais- ja Etelä-Suomen kanssa. Huomionarvoista oli, että pelkkien kevätkuukausien (huhti-toukokuu) sademäärä ei vielä vaikuttanut negatiivisesti lajin esiintymiseen. Tämä voisi viitata siihen, että sademäärä vaikuttaa juuri pesinnän onnistumiseen alkukesän sademäärän ollessa merkittävä tekijä: pelkän huhti-toukokuun sateet eivät välttämättä vaikuta voimakkaasti pesimäpaikan valikoitumiseen huolimatta muuton ajoittumisesta tälle ajalle.

Avoimen ja harvapuustoisien luonnonympäristön merkitys jäi lämpötilaa ja sademäärää pienemmäksi, mutta muuttujan lisääminen vaikutti mallien erottelukyvyn positiivisesti ilman, että erottelukyvyn hajonta olisi kasvanut merkittävästi mallien välillä. Piekanan vasteessa oli havaittavissa epäjohdonmukaisuutta, mikä viittaisi siihen, että käytetty muuttuja ei ilmennä riittävän hyvin piekanan habitaattivaatimuksia tai että avoimen luonnonympäristön kokonaismäärä ruudussa ei ole olennainen piekanan habitaatin valinnalle (Austin & Van Niel 2011). Tämä voi olla seurausta myös ruutujen koosta tai

yksilöiden sijoittumisesta ruudun sisällä (Guisan ym. 2007). Muuttujan käytön puolesta puhuu kuitenkin se, että mallit ennustivat suhteellisen esiintymistodennäköisyyden korkeimmaksi avoimen luonnonympäristön määrän ollessa lähellä maksimiarvoa (= 1 eli 100 % ruudusta). Tällaisia luonnontilaisia, avoimia habitaatteja on lähinnä Pohjois-Lapissa, jossa piekanahavaintoja oli tehty runsaasti. Ne tarjoavat piekanalle luontaisia saalistus- ja pesimäalueita ja toisaalta estävät metsistä riippuvaisten kilpailijalajien esiintymisen kyseisissä habitateissa (Potapov 1997).

Mehiläishaukankin mallinnuksessa käytetyt muuttujat ja niiden vaikutus lajin olivat ekologisesta näkökulmasta uskottavat. Lajin pohjoisimmat havainnot ovat Etelä-Lapista ja suurimmat keskittymät painottuvat aivan Etelä-Suomeen (kuva 16). Mehiläishaukan esiintymisalueen lämpötilat ja sadanta ovat korkeampia kuin pohjoisessa, joten lämpötilan positiivinen vaikutus lajin levinneisyyteen ja sadannan heikentynyt negatiivinen merkitys verrattuna piekanaa olivat kumpikin odotettuja. Lämpötilan merkitys korostui malleissa. Korkeissa lämpötiloissa muiden muuttujien arvoilla ei näyttänyt olevan suurta merkitystä lajin esiintymiselle. Koska mehiläishaukan ravinto koostuu erityisesti pesinnän aikaan lähes yksinomaan pistiäisistä, kuten ampiaisista (*Vespidae*), erityisesti niiden toukista ja muista hyönteisistä (Itämies & Mikkola 1972; Gensbøl ym. 1995), sopivan ravinnon määrän tulee olla riittävä. Kevään ja alkukesän lämpötilat vaikuttaa siihen, kuinka aikaisin ja kuinka paljon ampiaisia ja niiden toukkia on saatavilla (Lester ym. 2017), mikä näkyy mehiläishaukan pesinnän onnistumisessa ja populaatiokoossa (Kostrzewa 1989; Bijlsma ym. 2012). Viileinä vuosina mehiläishaukka joutuu todennäköisemmin turvautumaan vaihtoehtoiseen ravintoon, kuten lintuihin. Kuten aiemmin todettiin, lämpötila vaikuttaa positiivisesti kasvukauden aloitusajankohtaan, biomassan määrään ja esimerkiksi lehtipuiden osuuteen metsistä. Mehiläishaukka suosii reheviä, puulajistoltaan monipuolisia metsiä (Amcoff ym. 1994; Selås 1997), joita Suomessa tavataan eteläisessä Suomessa erityisesti ravinteikkailla kasvupaikoilla.

Sadannan kasvulla ei ollut yhtä voimakasta vaikutusta mehiläishaukkaan kuin piekanaa ja sademäärän merkitys jäi suhteessa huomattavasti lämpötilaa vähäisemmäksi. Runsaatkaan sademäärät eivät näyttäneet laskevan mehiläishaukan suhteellista esiintymistodennäköisyyttä lähelle nollaa alueilla, joilla lämpötila pysytteli riittävän korkeana. Vähäinen sademäärä oli kuitenkin lajin esiintymisen kannalta suotuisaa ja runsaat sateet laskevat suhteellista esiintymistodennäköisyyttä. Runsat sadanta voi vaikuttaa negatiivisesti poikasten selviytymiseen (Kostrzewa 1989), kuten aiemmin on havaittu piekanan ja muiden

päiväpetolintujen kohdalla. Sateinen kevät ja alkukesä voivat vaikuttaa negatiivisesti myös ampiaisten määrään (Tryjanowski ym. 2010; Lester ym. 2017). Eteläisen Suomen lämpimämmät säät voivat osaltaan lieventää sadannan haitallisia vaikutuksia mehiläishaukan esiintymiselle: runsaat sateet esimerkiksi ovat kohtalokkaimpia poikasille alhaisissa lämpötiloissa (Lehikoinen ym. 2009). Sademäärät vaikuttavat lämpötilan rinnalla positiivisesti kasvillisuuteen ja biomassan määrään ollen näin tärkeä paitsi mehiläishaukan suosimien rehevien metsien synnylle myös hyönteisille, jotka ovat riippuvaisia kasvillisuudesta (Altieri & Letourneau 1982). Mehiläishaukkojen pesät ovat piekanoita paremmassa suojassa sateelta, latvuserroksen alla (Kostrzewa 1998). Edellä mainitut seikat voivat olla yksi selittävä tekijä sille, miksi runsaatkaan sateet eivät välttämättä ole erityisen haitallisia mehiläishaukan esiintymiselle lämpimissä olosuhteissa.

Metsien määrän käyttö muuttujana paransi mallien erottelukykyä, vaikka muuttujan suhteellinen merkitys ei ollut suuri kummassakaan mallinnusvaiheessa. Mallien erottelukyvyn keskimääräinen hajontakaan ei näyttänyt lisääntyvän muuttujan lisäyksestä huolimatta. Mitä suurempi osuus tutkimusruudusta on metsää, sitä enemmän mehiläishaukalle pitäisi olla tarjolla sopivia habitaatteja (Amcoff ym. 1994; Selås 1997). Se, että metsien määrä ei vaikuttanut levinneisyyteen voimakkaammin, voi viitata muuttujan olevan liian yksinkertainen kuvaamaan lajille soveltuvia habitaatteja ainakin käytetyllä resoluutiolla. Suomessa metsät peittävät suuria järvi- ja asutusalueita sekä pohjoisinta Lappia lukuun ottamatta valtaosan maanpinnasta (ks. kuva 4c) ja metsien osuus keskimääräisestäkin ruudusta on suuri. Ruutujen arvoissa ei ole siis suurta hajontaa ja mehiläishaukkayksilöitä on luultavasti tavattu sekä kohtalaisesti että runsaasti metsää sisältävissä ruuduissa, minkä vuoksi muuttujan suhteellinen merkitys pysyttelee pienenä. Jatkossa kaikkien metsäalueiden määrää kuvaavan muuttujan sijaan voitaisiin hyödyntää useampaa muuttujaa, jotka kukin kuvaavat tietynlaisen metsän määrää: ovatko metsät lehtimetsiä, havumetsiä vai seka-metsiä, ovatko ne luonnontilaisia vai talousmetsiä ja niin edelleen. Vaikka Björklundin ynnä muiden (2015) tarkastelussa mehiläishaukan pesinnän onnistumisen kannalta metsän laadulla ei kuitenkaan vaikuttanut olevan merkitystä, lajin habitaatinvalinnassa metsän tyyppillä voi olla enemmän merkitystä, kuten esimerkiksi Selås (1997) esitti. Lassau ja Hochuli (2007) olivat havainnet, että ampiaisyhteisöt suosivat rakenteeltaan monipuolisia metsiä, joten metsien laatu ja määrä voivat vaikuttaa mehiläishaukan ravinnon saatavuuteen.

Toisessa mallinnusvaiheessa käytettiin muuttujana myös maatalousalueiden määrää ruuduissa. Mehiläishaukan vaste muuttujaan oli positiivinen ja muuttujan suhteellinen

merkitys oli jopa suurempi kuin sateen ja metsien määrän. Lajia onkin tavattu pesimässä maatalousalueiden ympäristöissä, erityisesti niittyjä ja avoimia heinikoita ympäröivissä metsissä (Kostrzewa 1998). Tällaisilla seuduilla pesimisessä voi olla lukuisia hyötyjä. Paitsi että tällaiset habitaatit ovat mehiläishaukan saalistamien hyönteislajien suosiossa, avoimet pellot ja niityt tarjoavat sopivaa metsästysmaastoa esimerkiksi lintujen saalistukseen hyönteisravinnon ollessa vähissä (Kostrzewa 1998. Kuten hiirihaukkojen, myös mehiläishaukan on havaittu välttelevän kanahaukan lähistöllä pesimistä (Gamauf ym. 2013). Kanahaukka välttää ihmisasutusta ja pyrkii pesimään vanhoissa, harventamattomissa havumetsissä (Gamauf ym. 2013; Björklund ym. 2015). Pesiminen muita päiväpetolintuja lähempänä ihmisasutusta ja maatalousalueita voi olla strategia vältellä kilpailua ja saalistajia sekä toisaalta mahdollisuus hyödyntää niiden ravintotarjontaa. Vesistöjen määrä ruudussa ei vaikuttanut olevan merkityksellinen, vaikka vesistöjen läheisyydellä on havaittu olevan merkitystä lajin pesimäpaikan valinnalle (Björklund ym. 2015). Vesistöjen määrä pesimäalueella ei ole välttämättä olennainen metsien petolinnulle, vaan merkitystä on enemmän juurikin etäisyydellä lähimpään vesistöön, kuten Björklund ynnä muut havaitsivat.

Havaintojen ja muuttujien suhde vaikuttaa mallien luotettavuuteen

Ennusteet voivat poiketa voimakkaasti todellisesta tilanteesta, jos havaintoihin, muuttujiin ja mallinnukseen heijastuvia epävarmuustekijöitä on riittävästi (Guisan & Thuiller 2005). Lopullisten GBM-mallien ennustuskyky esiintymishavaintoihin oli hyvä (Boyce indeksiksi), joten havaintoaineiston ja ympäristömuuttujien voidaan tulkita täyttäneen levinneisyysmallinnuksen vaatimuksen. Tämäkin tulkinta voi olla liian optimistinen sillä näennäisen hyviin ennusteisiin voidaan päätyä myös heikommalla aineistolla, jos käytetty menetelmä kykenee esimerkiksi yhdistämään epärealistiset havaintojen ja muuttujien suhteet toimivaksi malliksi (Austin & Van Niel 2011). Mallien esittämää suhdetta havaintojen ja muuttujien välillä voi olla vaikea tulkita ja arvioida: kaksiulotteiset vastekäyrät (Elith ym, 2005) esimerkiksi voivat käytännöllisyydestään huolimatta olla harhaanjohtavia useita muuttujia sisältävissä malleissa varsinkin pienillä aineistoilla (Zurell ym. 2012).

Jos lajien esiintymisten kartoitus ei ole ollut järjestelmällistä, on mahdotonta arvioida havaintojen maantieteellistä kattavuutta tai havaintojen oikeellisuuteen vaikuttavia tekijöitä (Graham ym. 2004; Phillips ym. 2009). Käytetty havaintoaineisto esimerkiksi on eittämättä painottunut helpommin saavutetuille alueille, kuten asutuksen lähelle, vaikka havaintoja onkin tehty eri puolilla Suomea, myös vaikeasti saavutettavilla seuduilla. Tämä

voi heikentää mallien ennusteita ja niiden luotettavuutta esimerkiksi vääristämällä lajin ja ympäristötekijöiden suhdetta (Guisan & Thuiller 2005; Mod ym. 2016). Vaikka esiintymishavainnoissa ei olisi virheitä, tausta-aineistoksi valikoituneissa tyhjiissä ruuduissa voi todellisuudessa olla yksilöitä, joita ei ole havaittu (Graham ym. 2004). Tyhjätkään ruudut eivät välttämättä ole huonoja habitaatteja lajille vaan jokin tuntematon tekijä on voinut estää esiintymisen niissä hetkellisesti (Soberón 2007).

Käytettyihin havaintoihin liittyy useita sellaisiakin epävarmuustekijöitä, joita ei ole edes mahdollista selvittää aineistoa yhdistäessä ja mallintaessa (Guisan & Thuiller 2005). Onko esiintymistä rajoittavan tekijän vaikutus habitaatissa pienempi johtuen esimerkiksi esiintymistä helpottavien tekijöiden positiivisesta vaikutuksesta? Onko alueella ollut havaintovuoden aikaan keskiarvoa paremmat olosuhteet lajin pesinnälle? Molemmat kysymykset ovat tärkeitä onnistuneen mallinnuksen kannalta, mutta niihin vastaaminen on ainakin tässä tapauksessa mahdotonta. Voimme kuitenkin arvioida tehtyjen valintojen vaikutusta edellä mainittujen kaltaisiin ongelmiin ja arvioida niiden riittävyttä suhteessa mallien ennustuskyykyyn. Havaintoaineisto ei ole täysin satunnaisesti kerätty vaan sen pohjana ovat Suomessa hyvin koordinoituneet linnustonseurannat, joita täydentävät yksittäisten lintuharrastajien tekemät havainnot (Valkama ym. 2011) ja valittuna olivat vain lajeista tehty pesimähavainnot. Aikavälin, jolla havainnot on tehty, tulisi olla riittävän pitkä, että havaintoja saadaan riittävän paljon eri puolilta lajin levinneisyysaluetta mutta samalla riittävän lyhyt, että samalla aikavälillä mitattujen ympäristömuuttujien vaikutus yksilöihin vastaisi todellista tilannetta luonnossa (Guisan ym. 2017: 140–141). Mallien perusteella näissä valinnoissa onnistuttiin kohtalaisen hyvin ja havaintojen määrä oli riittävä tuomaan esiin merkityksellisiä suhteita lajien ja ympäristön välillä. On myös todettava, että lintujen ominaisuudet, kuten erinomainen liikkumiskyky, voivat kompensoida aineistoon liittyviä epäkohtia kyetessään esimerkiksi seuraamaan ympäristöolosuhteiden vuosittaista vaihtelua.

Piekanan esiintymishavaintojen kohdalla oli pohdittava poikkeavien saalisvuosien vaikutusta mallinnukseen. Esimerkiksi vuosi 2009 oli erityisen hyvä myyrien, erityisesti peltomyyrien määrän suhteen, jolloin piekanoita ja muita harvinaisempia petolintulajeja pesi runsaasti Etelä- ja Keski-Suomessa (Lintuatlas kertoo pesimälinnuston... 2011). Aineistossa oli nähtävillä, että kyseisen vuoden lisäksi myös vuosina 2003, 2006 ja 2012 etelässä on havaittu useampia pesiviä pareja. Poikkeuksellisten vuosien havainnot olisi voitu poistaa ennen mallinnusta, millä puolestaan olisi voinut olla negatiivisiakin vaikutuksia mallien luotettavuuteen. Jos ruudut olivat pesimiseen soveltuvia, niiden poistaminen olisi

voinut johtaa muuttujien vaikutuksen yli- tai aliarvioimiseen malleissa (Mod ym. 2016) ja vähentänyt entisestään havaintojen määrää. Poistettavien havaintojen valinta olisi ollut vaikea tehdä objektiivisesti. Havaintojen säilyttämisen puolesta puhuivat myös käytetyt ympäristömuuttujat: myyrien esiintymistä, määrää tai niihin vaikuttavia tekijöitä ei mallinnettu mitenkään vaan esiintymisen mallinnus perustui ainoastaan ilmastoon ja avointen luonnonympäristöjen määrään. Vaikuttaakin siltä, että huolimatta esimerkiksi lämpötilan voimakkaasta merkityksestä lajin malleissa, ilmastotekijät eivät suoranaisesti estä piekanan esiintymistä etelämpänä vaan esiintymisalueen keskittyminen pohjoiseen on seurausta esimerkiksi bioottisista tai tuntemattomista tekijöistä.

Ylisovittaminen, tausta-aineisto ja spatiaalinen autokorrelaatio vaikuttavat ennusteisiin

Havaintojen ja muuttujien suhteeseen liittyvät ongelmat voivat johtaa herkästi mallien ylisovittamiseen (Holt ym. 2009; Merow ym. 2014). On oletettavaa, että ylisovittaminen aiheutti vaihtelua mallien välille huolimatta muuttujien rajaamisesta ja menetelmien keinoista rajoittaa mallien ylisovittamista. Ylisovittamista tapahtuu myös silloin, kun jokin taustalla oleva tuntematon tekijä vaikuttaa lajin esiintymiseen (*environmental noise*) tai kun otoksen jakauma on painottunut tai virheellinen (Hijmans 2012; Merow ym. 2014). Sattuma, dispersaali tai muut lajin kehityshistorialliset ilmiöt ja bioottiset suhteet ovat esimerkkejä lajiin vaikuttavista, ympäristötekijöiden vaikutuksen ulkopuolelle jäävistä tekijöistä, jotka voivat aiheuttaa malleihin epätodellisia vaikutuksia havaintojen ja muuttujien välille (Holt ym. 2009; Boulangeat ym. 2012). Mitä monimutkaisempia menetelmät ja mallinnettavat suhteet havaintojen ja muuttujien välillä ovat ja mitä enemmän parametreja on käytännössä, sitä todennäköisemmin mallit altistuvat ylisovittamiseen (Merow ym. 2014). Ylisovittamisen riskiä on voinut lisätä lajien levinneisyyteen vähemmän vaikuttaneiden maanpeitemuuttujien lisääminen malliin (Dormann ym. 2013).

Havaintoaineistoa olisi voinut jakaa osiin, jotta mallien suorituskykyä ja luotettavuutta voitaisiin arvioida objektiivisemmin. Mallien ennustuskyky olisi voinut laskea käytössä olevien havaintojen määrän vähentyessä, mutta ylisovittamisen kaltaiset epävarmuustekijät olisivat todennäköisesti selvinneet vertaamalla sovittamiseen käytetyn aineiston ja siitä erillisen evaluaatioaineiston erottelukyvyn arvoja (Franklin 2009; Guisan ym. 2017). Parhaiten ylisovittamista kyetään kuitenkin arvioimaan, kun mallien ennustuskykyä käytetyssä aineistossa verrataan saman mallin ennustuskykyyn täysin itsenäisellä aineistolla

(Peterson ym. 2011). Näissä malleissa käytetyn aineiston jakaminen suoraan mallien sovitamiseen ja evaluaatioon käytettäviin otoksiin ei tuottaisi täysin itsenäistä aineistoa (Hijmans 2012). Molemmissa otoksissa olisi nähtävissä esimerkiksi samat aineistonkeruuseen liittyvät valinnat ja tuntemattomien tekijöiden vaikutus havaintojen jakautumiseen ympäristössä eikä otoksia näin voitaisi pitää itsenäisinä (Peterson ym. 2011). Vertailuaineistona olisikin hyvä käyttää eri ajanjaksolta tai alueelta saatua aineistoa, joka ei sekään välttämättä täytä kaikkia itsenäisen aineiston piirteitä (Randin ym. 2006; Hijmans 2012). Myös otoskoko varsinkin piekanan kohdalla voisi rajautua liian pieneksi, jos evaluaatioaineisto pyritäisiin pitämään täysin erillään mallien sovittamisessa käytetystä aineistosta.

Ylisovittaminen ei välttämättä ole niin suuri uhka ennustettaessa levinneisyyttä tutkimusalueen sisällä (Merow ym. 2014). Ylisovittamisen korjaaminen välttämätöntä, jos malleja haluttaisiin yleistää tai siirtää uusille alueille tai uuteen aikaan (Randin ym. 2006; Pearman ym. 2008). Ylisovittamisen tunnistaminen ja kontrollointi olisi tärkeää ennustamisessakin, koska liian voimakkaat, ekologisesti merkityksettömät suhteet havaintojen ja muuttujien välillä voivat esimerkiksi johtaa tilanteeseen, jossa lajin kannalta riittävän hyviä habitaatteja sivuutetaan huonoina ylisovittamisen seurauksena (Randin ym. 2006; Merow ym. 2014). Tämän todennäköisyys kasvaa esimerkiksi silloin, kun aineistossa on ruutuja, joissa lajin esiintyminen tai puuttuminen on kirjattu virheellisesti.

Vaihtelua voi aiheuttaa myös tausta-aineistojen määrä ja jakautuminen otoksissa (Phillips ym. 2009; Barbet-Massin ym. 2012). Mallinnusvaiheissa havaittiin, että tausta-aineiston määrän muutos vaikutti erottelukykyyn ja sen vaihteluun ja lopulliset määrät pyrittiin valitsemaan siten, että mallien ennustuskky ja sen vaihtelu olivat tasapainossa. Tausta-aineiston valikoitumisessa tehdyt valinnat ja rajaukset saattoivat vahvistaa aineistossa olevia epävarmuustekijöitä ja tuottaa näin epäluotettavia tai vääriä malleja huolimatta niiden hyvästä ennustuskvyyistä (Wisz & Guisan 2009). Tausta-aineiston valikoitumista olisikin voitu ohjata erilaisilla otantamenetelmillä mallien parantamiseksi. Tausta-aineisto olisi voitu ottaa esimerkiksi ruuduista, joista oli havaittu muita päiväpetolintulajeja (Phillips ym. 2009) tai ruuduista, joissa kyseisiä lajeja on havaittu aiemmissa kartoituksissa (Lütolf ym. 2006). Tällaisetkaan kriteerit malleihin valikoituvalle tausta-aineistolle eivät välttämättä ole luotettavuudeltaan sen parempia kuin satunnaisotanta (Lütolf ym. 2006; Phillips ym. 2009). Optimaalinen tausta-aineiston valinta voi olla hyvinkin tutkimus- ja menetelmäkohtaista (Barbet-Massin ym. 2012). Mallit kykenivät ennustamaan kohtalaisen hyvin lajien esiintymiset, joten voidaan olettaa, että tausta-aineisto valikoitui riittävän

hyvin edustamaan alueita, jotka ovat lajille epäsopivia; tausta-aineiston rajausta oli kuitenkin yksi subjektiivisimmista valinnoista mallinnoissa ja siten suuri riski mallien luotettavuuden kannalta (Wisz & Guisan 2009; Phillips ym. 2009).

Lopullisten mallien residuaaleissa havaittiin spatiaalista autokorrelaatiota. Itse havaintojen välillä olevaa riippuvuutta ei mitattu, mutta on oletettavaa, että niiden jakautuminen vaikutti havaittuun riippuvuuteen residuaaleissa. Spatiaalinen autokorrelaatio voi johtua biologisesta prosessista tai jostain tuntemattomasta tekijästä, joka ohjaa lajien yksilöitä samalla alueella, kuten ilmastollinen tai bioottinen tekijä (Guisan & Thuiller 2005). Yksinkertaisia tapoja vähentää havaintojen spatiaalista autokorrelaatiota ovat lähekkäisten havaintojen poisto tai otoksen havainnoille säädetty vähimmäisetäisyys (Dormann ym. 2007), mitkä kuitenkin pienentäisivät otoskokoa ja mahdollisesti heikentäisivät näin mallien ennustuskkyä ja luotettavuutta. Havaintojen keskittymät voivat hyvinkin olla seurausta todellisista eroista habitaattien välillä (Albert ym. 2010). Linnut esimerkiksi kykenevät asuttamaan useimmat alueet erinomaisen dispersaaliokykynsä ansiosta, joten ajan puute, häiriötekijät tai muut vastaavat tekijät eivät tehokkaasti rajoita yksilöiden asuttamien habitaattien määrää tai sijaintia (Barve ym. 2011). Havaintojen suuren ruutukoon myötä läheisten ruutujen yksilöt ovat todellisuudessa kohtalaisen etäisyyden päästä toisistaan ja ruudut kattavat heterogeenisiä habitaatteja. Lajeihin vaikuttavia tekijöitä ei mallinnettu, mutta esimerkiksi saalislajien esiintyminen voi ohjata yksilöiden esiintymistä eri puolilla Suomea luoden painotuksia yksilöiden levinneisyydessä (Boulangeat ym. 2012).

Spatiaalista autokorrelaatiota tulisi pyrkiä mahdollisuuksien mukaan vähentämään myös levinneisyysmallinnuksen kaltaisissa tilastotieteeseen perustuvissa tutkimusmenetelmissä huolimatta spatiaaliseen autokorrelaatioon usein johtavista luonnollisista tekijöistä (Thibaud ym. 2014). Vielä tärkeämpää olisi saada nämä mukaan mallinnukseen joko muutujina tai muilla keinoin, sillä ne voivat parantaa mallien ennustuskkyä ja ennen kaikkea niiden luotettavuutta (Thibaud ym. 2014; Norberg ym. 2019). Spatiaalinen autokorrelaatio ei välttämättä ole niin merkittävä tekijä kuin levinneisyysmallinnuksen parissa työskennelleet ovat epäilleet, kuten Thibaud ynnä muut (2014) totesivat tarkastellessaan eri epävarmuustekijöiden vaikutusta levinneisyysmallinnusten onnistumiseen ja luotettavuuteen. Spatiaalisen autokorrelaation ilmeneminen on joka tapauksessa negatiivista mallien luotettavuuden näkökulmasta ja heikentää niiden uskottavuutta.

Aineiston rajoitteet eivät estäneet lajien levinneisyysmallinnusta

Tulosten tarkastelun pohjalta voidaan todeta, että pelkkien esiintymishavaintojen käyttäminen ja aineiston matala resoluutio eivät estä hyvän ennustuskyvyn omaavien levinneisyysmallien luomista. Aineiston epävarmuustekijät kuitenkin heijastuvat tulosten luotettavuuteen ja mallien hyvään ennustuskykyyneen tulee suhtautua kriittisesti ja varauksella (Cheafaoui & Lobo 2008; Phillips ym. 2009; Austin & Van Niel 2011). Kaikissa mallinnukseen liittyvissä valinnoissa ei kyetty riittävään objektiivisuuteen aineiston rajoitusten vuoksi ja tavoiteltaessa esiintymistä mahdollisimman hyviä ennustavia malleja.

Koska tiedossa oli, että havaintojen luotettavuuteen liittyy epävarmuustekijöitä, niiden valikoitumista mallinnukseen olisi ollut järkevää rajoittaa erilaisilla satunnaisotantamenetelmillä (Albert ym. 2010), mikä olisi tapahtunut havaintojen määrän ja ennustuskyvyn kustannuksella. Kuten aiemmissa tarkasteluissa oli todettu (McPherson ym. 2006; Guisan ym. 2007; Luoto ym. 2007), suuri ruutukoko ei välttämättä estä levinneisyyden mallintamista suurella tutkimusalueella. Päiväpetolinnut kykenevät liikkumaan suurella alueella, joten liian pieni ruutukoko voisi johtaa samojen yksilöiden laskemiseen useampaan kertaan (Thibaud ym. 2014). Toisaalta yksilöiden liikkumiskyky aiheuttaa ongelmia, sillä ne voivat liikkua myös useampien kuin yhden ruudun alueella, mikä lisää malleihin liittyvää epävarmuutta (Björklund ym. 2015). Suuri ruutukoko voi kuitenkin kasvattaa ylisovittamisen riskiä sillä vaarana on, että lajin ja ympäristön suhteen sijaan mallinnetaan habitaatinvalintaan liittymättömiä piileviä tekijöitä (McPherson ym. 2006; Guisan ym. 2007; Seo ym. 2009). Mallien esittämää muuttujan ja havainnon suhdetta ei siis tule ylitulkitä.

Lajien vasteet ympäristömuuttujiin olivat odotetun kaltaisia ja riittävän yksinkertaisia. Mallien perusteella näytti, että kylmyys ei estä piekanan esiintymistä sillä lajin suhteellinen esiintymistodennäköisyys oli suurimmillaan minimilämpötilassa. Suomi sijaitsee itse asiassa piekanan levinneisyysalueen etelärajalla, joten sopeutuminen kylmään ympäristöön ei tule yllätyksenä (The IUCN Red List... 2020). Mehiläishaukka puolestaan vaatii korkeita lämpötiloja, mikä näkyy myös sen habitaatinvalinnassa ja ravinnossa (Kostrzeva 1989, 1997). Sademäärä vaikuttaa negatiivisesti molempien esiintymiseen, todennäköisesti pesinnän onnistumisen ja ravinnon saatavuuden myötä (Kostrzeva 1989; Terraube ym. 2015). Nimenomaan muuton ja pesinnän aikaiset ilmasto-olosuhteet näyttäisivät olevan lajien kannalta tärkeitä.

Vaikka maanpeitemuuttajat paransivat hieman mallien ennustuskyykyä, niiden merkitys oli pieni levinneisyysalueen muodostumisessa verrattuna ilmastoon. Vastaavia muuttujia on aieminkin käytetty onnistuneesti yhdistettynä ilmastomuuttujiin (Virkkala ym. 2005; Heikkinen ym. 2007; Luoto ym. 2007). Käytetty CORINE-aineisto ei välttämättä ollut riittävä kuvaamaan lajien suosimia habitaatteja. On myös mahdollista, että suuri ruutukoko peitti habitaattien vaihtelua ja että habitaatin määrän keskiarvo ruudussa ei ollut hyvä mittari erottamaan erilaisten habitaattien laatua lajille (Guisan ym. 2007). Monipuolisesti habitaattien laatua kuvaavat muuttajat habitaatin ääriarvoja tai vaihtelua ja havainnon ja habitaatin etäisyyksiä kuvaavien muuttujien käyttö voisi paremmin kuvata lajin ja habitaatin suhdetta varsinkin, jos aineiston resoluutiota voitaisiin kasvattaa (mm. Warren ym. 2005; Titeux ym. 2007; Mattsson ym. 2013; Shirley ym. 2013). Habitaattien määrää ja jakautumista voitaisiin tukea lisäämällä niihin habitaattien laatua kuvaavia arvoja, esimerkiksi kaukokartoitusmenetelmien avulla (He ym. 2015). Tutkimusasetelma ja havaintoaineisto kuitenkin puolsivat yksinkertaisen keskiarvon käyttöä habitaatin määrän kuvaajana. Käytettyjen CORINE-luokkien sijaan taas oltaisiin voitu käyttää myös ylimääraisiä tai vaihtoehtoisia luokkia (Wisz ym. 2008a). Erilaiset muuttujan arvot olisivat voineet parantaa myös ilmastomuuttujien käyttöä. Sademäärän keskiarvoa tärkeämpää päiväpetolintujen esiintymiselle voivat olla esimerkiksi äärimmäisen sateisten keväiden ja kesien tai jopa yksittäisten kovasateisten päivien määrä alueittain, koska äkilliset rankkasateet vaikuttavat olevan esimerkiksi poikasten selviytymisen kannalta erityisen vaarallisia (Mearns & Newton 1988; Robinson ym. 2017).

Tärkeä keino vähentää havaintojen ja muuttujien suhteen ulkopuolista vaihtelua malleissa olisi saada bioottisen ympäristön vaikutus käytettävien muuttujien joukkoon (Kissling ym. 2012; Pollock ym. 2014; Anderson ym. 2017; Dormann ym. 2018). Bioottisen ympäristön mallintaminen voi paljastaa mielenkiintoisia ja aiemmin tuntemattomia vuorovaikutuksia lajien välillä (Dormann ym. 2018). Pokrovsky ynnä muut (2020) ja Seyer ynnä muut (2020) esimerkiksi löysivät arktisilla alueilla pesivien lintujen välillä kiinnostavia, lajien habitaatinvalintaan vaikuttavia käyttäytymismalleja. Piekana kuului molemmissa tutkimuksissa tarkasteltuihin lajeihin. Havaintojen maantieteellisessä jakautumisessa näkyy kuitenkin jo bioottisten vuorovaikutussuhteiden vaikutus lajeihin, joten osa päiväpetolintujen saalislajien esiintymisen tai lajinsisäisen ja -välisen kilpailun vaikutuksista välittyy myös malleihin (Wisz ym. 2013; Anderson ym. 2017). Mallit voivat liittää näitä vaikutuksia virheellisesti esimerkiksi lajien ja ilmastotekijöiden väliseen suhteeseen ja esittää

lajien vasteet väärin (Kissling ym. 2012; Pollock ym. 2014). On myös muistettava, että eliöyhteisöissä jokainen laji altistuu samojen ympäristötekijöiden vaikutukselle, mutta lajien vasteet niihin eroavat toisistaan (Whittaker 1967). Toisaalta havaitun esiintymisen taustalla voi olla myös jokin tuntematon tekijä, joka ohjaa eri lajien levinneisyyttä samaan suuntaan ja välittyy malliin esimerkiksi ilmaston kautta (Austin 2002).

Päiväpetolintujen levinneisyyttä voidaan ennustaa useilla menetelmillä

Tutkielmassa käytetty, melko yksinkertainen tapa vertailla menetelmien suorituskkyä ei mahdollista absoluuttisesti parhaan menetelmän valintaa kullekin lajille. Vertailu kuitenkin auttoi rajaamaan menetelmävaihtoehtojen joukosta pois valittujen lajien levinneisyyttä heikosti ennustavat menetelmät ja korostamaan ne menetelmät, joita voidaan käyttää habitaattivaatimuksiltaan, ravinnoltaan ja maantieteelliseltä levinneisyydeltään poikkeavien päiväpetolintulajien levinneisyyden mallintamisessa. Tarkempi vertailu olisi mahdollista luomalla yksittäiset, huolella sovitettut mallit myös GLM-, GAM-, ja RF-menetelmillä. Koska regressiomenetelmät ovat tietokoneoppimisen menetelmiin verrattuna hyvin menetelmälähtöisiä, niiden käyttö on suositeltavaa erityisesti tilanteissa, joissa käytössä oleva aineisto on luotettava ja kattava (Elith & Leathwick 2009). GBM ja RF ovat joustavia ja kykenevät mallintamaan hyvin monimutkaisiakin lajin ja ympäristön välisiä suhteita, mikä kuitenkin altistaa mallit ylisovittamiselle ja näin todellisesta levinneisyydestä poikkeaville ennusteille (Phillips ym. 2009; Barbet-Massin ym. 2012; Merow ym. 2014). GLM ja GAM ovat siten ehkä turvallisempi lähtökohta mallinnukseen ja tarjoavat helpommin tulkittavia työkaluja mallien sovittamiseen ja evaluaatioon (Elith & Leathwick 2009; Merow ym. 2014). Se, että lajien vasteet käytettyihin muuttujiin eivät olleet erityisen monimutkaisia tai hierarkisia, voi olla yksi syy sille, miksi regressiomenetelmät tuottivat erottelukyvyltään GBM:n ja RF:n veroisia malleja (Merow ym. 2014).

Huolimatta havaitusta yhtenäisyydestä eri menetelmien ennustuskkyssä voidaan todeta, että vertailu oli onnistunut sen auttaessa rajaamaan huonon erottelukyvyn menetelmät päiväpetolintujen levinneisyysmallinnuksesta. Regressiomenetelmien, GBM:n ja RF:n soveltuminen lajien mallinnukseen antoi myös uskoa siihen, että lajien ja ympäristön suhde kyettiin mallintamaan riittävän luotettavasti. Levinneisyysmallinnuksessa on käytetty myös lukuisia tämän vertailun ulkopuolisia menetelmiä, joita voitaisiin hyödyntää myös päiväpetolintujen levinneisyysmallinnuksessa (mm. Franklin 2009, Guisan ym. 2017; Norberg ym. 2019). Norbergin ynnä muiden (2019) kattavassa vertailussa GLM, GAM ja GBM

(BRT) olivat suorituskyvyltään keskivertoa ja RF oli kautta linjan heikoimpien menetelmien joukossa. He painottivat aineiston luonteen ja lajin ominaisuuksien vaikutusta menetelmien suorituskyvyn ja suosittelivat pohtimaan tarkemmin myös itse ennustuskyvyn arvioinnissa käytettyjen kriteerien valintaa tutkimuskohtaisesti.

Menetelmien vertailussa ja mallien luotettavuuden arvioinnissa käytettyjen arviointikriteerien sijaan oltaisiinkin voitu käyttää myös vaihtoehtoisia ennustuskyvyn mittareita (Norberg ym. 2019; Wunderlich ym. 2019). Boycen indeksin laskeminen on tällä hetkellä käyttökelpoisimpia tapoja arvioida pelkkiä esiintymishavaintoja hyödyntävien mallien ennustuskkyä (Di Cola ym. 2017), joten sen käyttö oli perusteltua. Esimerkiksi mehiläishaukan malleissa havaittu matala spesifisyys laski TSS- ja AUC-arvoilla arvioitua mallien ennustuskkyä lajille, vaikka mallin ennustuskkyä lajiyksilöistä ”tyhjiin” ruutuihin (tausta-aineisto) ei sinänsä voitu mitata. TSS:n ja AUC:n luotettavuus on todistetusti riippuvainen myös aineiston koosta ja havaintojen ja tausta-aineistojen suhteellisesta jakaumasta otoksessa (ns. *prevalence*, McPherson ym. 2004; Lobo ym. 2008; Guisan ym. 2017: 119–120). Huolimatta tästä kritiikistä on todettava, että monivuotiset linnustonseurannat ovat olleet Suomessa sen verran kattavia, että luultavasti suuri osa tausta-aineistoruuduista on todellakin sellaisia, joissa yksilöitä ei esiinny. Näin ollen mallien ennustuskky voi todellisuudessa ollakin esimerkiksi TSS-arvon mukainen eli heikompi kuin mitä Boycen indeksin mittaama ennustuskky. Ennustuskvyn tueksi eri mallien ennusteiden vaihtelua tutkimusruuduissa olisi voitu tarkastella erikseen. Tämä olisi onnistunut esimerkiksi luomalla karttaesitys, joka paljastaisi alueet, joilla mallien ennusteissa oli eniten vaihtelua (*uncertainty maps*, mm. Argáez ym. 2005).

Käytettyä aineistoa voidaan tukea ja hyödyntää monin eri tavoin

Muuttujien tai aineiston muokkaamisen ja algoritmin valinnan lisäksi voidaan esittää myös muita keinoja parantaa luotujen mallien ennustuskkyä ja luotettavuutta. Erilaiset menetelmät ja lajien ekologian vaihtoehtona yksittäisen menetelmän käytölle mallien sovittamisessa olisi hyödyntää eri menetelmien vahvuuksia niin sanotussa *ensemble*-mallinnuksessa (Thuiller ym. 2009). *Ensemble*-mallit yhdistävät tai yleistävät eri menetelmien mallit yhdeksi ennusteeksi, jonka tulisi olla tarkempi kuin yksittäisen mallin ennuste ja parantaa samalla levinneisyysmallinnuksen luotettavuutta (Araújo & New 2007). Näin voitaisiin hyödyntää esimerkiksi tässä tutkielmassa käytettyjen menetelmien vahvuudet ja luoda yksittäistä menetelmää parempia malleja (Thuiller ym. 2009). BIOMOD2-paketti esimerkiksi

mahdollistaa eri menetelmien ennusteiden yhdistämisen melko yksinkertaisesti. *Ensemble*-menetelmien käyttö voi kasvaa tulevaisuudessa, kun niiden käytöstä ja suorituskyvystä verrattuna tavallisiin levinneisyysmalleihin saadaan enemmän tutkimusnäyttöä (Hao ym. 2019).

Bioottisten vuorovaikutusten tuominen levinneisyysmalleihin muuttujina on haastavaa johtuen esimerkiksi niiden monimutkaisesta vaikutuksesta lajien levinneisyyteen verrattuna abioottisiin tekijöihin (Kissling ym. 2012; Wisz ym. 2013; Anderson 2017; Dormann ym. 2018). Viime vuosina onkin siirrytty kohti menetelmiä, joissa mallinnetaan yhtä aikaa useita eri lajeja: näitä kutsutaan yhteisnimellä *joint-species distribution models* (Pollock ym. 2014). Niissä muiden lajien esiintymistä käytetään selittämään sitä vaihtelua levinneisyydessä, jota abioottisilla tekijöillä ei ole pystytty selittämään, ei siis varsinaisina selittävinä muuttujina (Kissling ym. 2012; Dormann ym. 2018). Näin voitaisiin löytää mahdollisia vuorovaikutuksia lajien välillä ja parantaa mallien suorituskkyä varsinkin tilanteissa, joissa tuntemus lajien välisistä suhteista on puutteellista. On kuitenkin mahdollista, että näissäkkin malleissa havaittuja korrelaatioita ohjaavat esimerkiksi yhteisten habitaattien suosiminen, lajien historia alueella ja malleista puuttuvat ympäristötekijät (Dormann ym. 2018). Levinneisyysmallinnuksen parissa tulisi päästä jatkossa sellaiseen tilanteeseen, jossa bioottinen ympäristö esitettäisiin malleissa luotettavasti, todellisten vaikutusten kautta (Wisz ym. 2013; Pollock ym. 2014; Dormann ym. 2018).

Levinneisyysmallinnuksen tueksi tulisi mahdollisuuksien mukaan lisätä myös muunlaista tietoa lajin kannalta olennaisista ekologisista prosesseista, esimerkiksi populaatiodynamiikasta, mikä voisi esimerkiksi johdatella mallinnuksessa käytettävien havaintojen valintaa sellaisiin habitaatteihin, jotka ovat lajin kannalta varmasti suotuisia (Pulliam 2000; Chase & Leibold 2003; Borregaard ym. 2010). Erilaisia ekologisia malleja voidaan myös yhdistää levinneisyysmalleihin (Guisan ym. 2017; Miller ym. 2019). Useimpien lajien levinneisyysmallinnuksessa tulisi huomioida esimerkiksi lajin dispersaaliokyky, joka väistämättä rajoittaa lintuja rajoittuneempien lajien levinneisyysaluetta (Barve ym. 2011; Peterson ym. 2011; Boulangeat ym. 2012), tai lajien yksilöllinen kyky sietää erilaisia abioottisia ja bioottisia tekijöitä ja niiden muutoksia (Austin 2007). Toisaalta itse havaintoaineistoa voitaisiin käsitellä siten, että vastemuuttujana olisi yksilöiden määrä ruuduissa tai tietyn suuruudessa ympäristössä – tällaisia ovat esimerkiksi *point process* -mallit, joita Renner ynnä muut (2015) suosittelevat *presence only* -aineiston avulla suoritettavaan levinneisyysmallinnukseen.

Johtopäätökset

Päiväpetolintujen levinneisyysmallinnus tuotti aineiston rajoitteista huolimatta hyviä ennusteita piekanan ja mehiläishaukan esiintymisestä Suomessa. Myös muiden päiväpetolintujen levinneisyyttä voitaisiin oletettavasti mallintaa vastaavalla aineistolla ja menetelmillä. Menetelmien vertailusta oli hyötyä mallien uskottavuuden tulkinnassa ja lajeille parhaiten soveltuvat menetelmät onnistuttiin erottamaan huonosti suoriutuvista menetelmistä. Levinneisyysmallinnus auttoi kartoittamaan lajien levinneisyyttä Suomessa ja arvioimaan lajeihin vaikuttavien tekijöiden voimakkuutta ja suhdetta sekä asettamaan suuntaviivoja päiväpetolintujen levinneisyysmallinnukselle jatkossa. Edellä mainittujen kaltaiset havainnot voivat auttaa selvittämään lajien uhanalaisuuden taustalla olevia syitä ja tukemaan lajien ja lajeille tärkeiden habitaattien suojelua. Päiväpetolintujen suojelu voi tukea myös muiden, samoissa habitaateissa esiintyvien lajien ja itse ekosysteemien suojelua. Mallien parantaminen on mahdollista esimerkiksi uusien ja paremmin ympäristömuuttujien, tausta-aineiston valintaa tarkentamalla ja uusia vaihtoehtoisia mallinnustapoja hyödyntäen.

Ennusteita **ei kuitenkaan tule ekstrapoloida** uusille alueille tai uuteen aikaan ja mallinnuksessa käytettyjen ympäristötekijöiden ja lajien suhdetta tulee pitää osittain suuntaa antavana. Ennusteiden luotettavuuteen liittyy epävarmuutta, jonka korjaaminen tulisi asettaa etusijalle esimerkiksi mallinnettaessa uusien päiväpetolintulajien levinneisyyttä. Niiden luotettavuutta voi halutessaan parantaa keskittymällä tarkemmin levinneisyysmallinnuksen varrella havaittuihin epävarmuustekijöihin. Esimerkiksi spatiaalista autokorrelaatiota tulisi vähentää huomattavasti. Tutkimuksen tavoite määrittelee lopulta, pyritäänkö mallinnuksessa maksimoimaan mallin ennustuskyky vai voiko ennustuskyvyssä joustaa ennusteiden yleistettävyyden ja luotettavuuden varmistamiseksi. On kuitenkin tärkeää varmistaa, että samalla kun mallinnuksissa pyritään mahdollisimman hyvään ennustuskykyyn, myös epävarmuustekijöiden minimoimiseen kiinnitetään huomiota luotettavien ja käyttökelpoisten ennusteiden saavuttamiseksi. Vain näin voidaan varmistaa, että levinneisyysmallinnuksen avulla saatuja tuloksia voidaan hyödyntää todellisten ongelmien ratkaisussa. Lajien levinneisyyden ennustaminen ja riittäviä populaatiokokoa ylläpitämään kykenevien habitaattien paikantaminen ovat tärkeitä lajien, lajiyhteisöjen ja ekosysteemien suojelun näkökulmasta jokaisella maantieteellisellä tarkastelutasolla. Levinneisyysmallinnus ja sen kehittyminen mahdollistavat ilmastomuutoksen ja habitaattikadon vaikutusten tunnistamisen ja auttavat koordinoimaan suojelutoimenpiteiden aloittamista ja tehostamista.

Kirjallisuus

- Albert, C. H., N. G. Yoccoz, T. C. Edwards Jr, C. H. Graham, N. E. Zimmermann & W. Thuiller (2010). Sampling in ecology and evolution—bridging the gap between theory and practice. *Ecography* 33: 1028-1037.
- Allouche, O., A. Tsoar & R. Kadmon (2006). Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS). *Journal of Applied Ecology* 43: 1223-1232.
- Altieri, M. A. & D. K. Letourneau (1982). Vegetation management and biological control in agroecosystems. *Crop Protection* 1: 405-430.
- Amcoff, M., M. Tjernberg & Å. Berg (1994). Bivråkens *Pernis apivorus* boplatsval. *Ornis Svecica* 4: 145-158.
- Anttil, A., A. Franke & J. Bêty (2014). Heavy rainfall increases nestling mortality of an arctic top predator: experimental evidence and long-term trend in peregrine falcons. *Oecologia* 174: 1033-1043.
- Anderson, R. P. (2017). When and how should biotic interactions be considered in models of species niches and distributions? *Journal of Biogeography* 44: 8-17.
- Araújo, M. B. & M. Luoto (2007). The importance of biotic interactions for modelling species distributions under climate change. *Global Ecology and Biogeography* 16: 743-753.
- Araújo, M. B. & A. Rozenfeld (2014). The geographic scaling of biotic interactions. *Ecography* 37: 406-415.
- Araújo, M. B., R. G. Pearson, W. Thuiller & M. Erhard (2005). Validation of species–climate impact models under climate change. *Global Change Biology* 11: 1504-1513.
- Araújo, M. B. & A. Guisan (2006). Five (or so) challenges for species distribution modelling. *Journal of Biogeography* 33: 1677-1688.
- Araújo, M. B. & A. T. Peterson (2012). Uses and misuses of bioclimatic envelope modeling. *Ecology* 93: 1527-1539.
- Araújo, M. B. & M. New (2007). Ensemble forecasting of species distributions. *Trends in Ecology & Evolution* 22: 42-47.
- Argáez, J., J., Andrés, M., Nakamura & J., Soberón. (2005). Prediction of potential areas of species distributions based on presence-only data. *Environmental and Ecology Statistics* 12, 27-44.
- Austin, M. (2007). Species distribution models and ecological theory: A critical assessment and some possible new approaches. *Ecological Modelling* 200: 1-19.
- Austin, M. P. (1990). Community theory and competition in vegetation. *Perspectives on plant competition*: 215-238.
- Austin, M. P. & T. M. Smith (1989). A new model for the continuum concept. *Vegetatio*: 35-47.
- Austin, M. P. & K. P. Van Niel (2011). Improving species distribution models for climate change studies: variable selection and scale. *Journal of Biogeography* 38: 1-8.
- Barbet-Massin, M., F. Jiguet, C. H. Albert & W. Thuiller (2012). Selecting pseudo-absences for species distribution models: How, where and how many? *Methods in Ecology and Evolution* 3: 327-338.
- Barry, S. & J. Elith (2006). Error and uncertainty in habitat models. *Journal of Applied Ecology* 43: 413-423.
- Barve, N., V. Barve, A. Jiménez-Valverde, A. Lira-Noriega, S. P. Maher, A. T. Peterson, J. Soberón & F. Villalobos (2011). The crucial role of the accessible area in ecological niche modeling and species distribution modeling. *Ecological Modelling* 222: 1810-1819.
- Bijlsma, R., M. Vermeulen, L. Hemerik & C. Klok (2012). Demography of European Honey Buzzards *Pernis apivorus*. *Ardea* 100: 163-177.
- Björklund, H., J. Valkama, E. Tomppo & T. Laaksonen (2015). Habitat effects on the breeding performance of three forest-dwelling hawks. *PLoS one* 10: e0137877.
- Björklund, H., A. Santangeli, F. G. Blanchet, O. Huitu, H. Lehtoranta, H. Lindén, J. Valkama & T. Laaksonen (2016). Intraguild predation and competition impacts on a subordinate predator. *Oecologia* 181: 257-269.
- Borregaard, M. & C. Rahbek (2010). Causality of the Relationship between Geographic Distribution and Species Abundance. *The Quarterly review of biology* 85: 3-25.

- Boucher, D. H., S. James & K. H. Keeler (1982). The Ecology of Mutualism. *Annual Review of Ecology and Systematics* 13: 315-347.
- Boulangeat, I., D. Gravel & W. Thuiller (2012). Accounting for dispersal and biotic interactions to disentangle the drivers of species distributions and their abundances. *Ecology Letters* 15: 584-593.
- Boyce, M. S., P. R. Vernier, S. E. Nielsen & F. K. A. Schmiegelow (2002). Evaluating resource selection functions. *Ecological Modelling* 157: 281-300.
- Brown, J., G. Stevens & D. Kaufman (1996). The geographic range: Size, shape, boundaries, and internal structure. *Annual Review of Ecology and Systematics* 27: 597-623.
- Bruno, J. F., J. J. Stachowicz & M. D. Bertness (2003). Inclusion of facilitation into ecological theory. *Trends in Ecology & Evolution* 18: 119-125.
- Byholm, P., S. Mäkeläinen, A. Santangeli & D. Goulson (2018). First evidence of neonicotinoid residues in a long-distance migratory raptor, the European honey buzzard (*Pernis apivorus*). *Science of The Total Environment* 639: 929-933.
- Carothers, J. H. & F. M. Jaksic (1984). Time as a Niche Difference: The Role of Interference Competition. *Oikos* 42: 403-406.
- Chase, J. M. & M. A. Leibold (2003). *Ecological niches: Linking classical and contemporary approaches*. 221 s. University of Chicago Press, Chicago.
- Chefaoui, R. & J. Lobo (2008). Assessing the effects of pseudo-absences on predictive distribution model performance. *Ecological Modelling* 4: 478-486.
- Chen, I.-C., J. K. Hill, R. Ohlemüller, D. B. Roy & C. D. Thomas (2011). Rapid range shifts of species associated with high levels of climate warming. *Science* 333: 1024-1026.
- Connell, J. H. (1961). The Influence of Interspecific Competition and Other Factors on the Distribution of the Barnacle *Chthamalus Stellatus*. *Ecology* 42: 710-723.
- Cord, A. & D. Rödder (2011). Inclusion of habitat availability in species distribution models through multi-temporal remote-sensing data? *Ecological Applications* 21: 3285-3298.
- Cutler, D. R., T. C. Edwards Jr., K. H. Beard, A. Cutler, K. T. Hess, J. Gibson & J. J. Lawler (2007). Random Forests for classification in ecology. *Ecology* 88: 2783-2792.
- Deblauwe, V., V. Droissart, R. Bose, B. Sonké, A. Blach-Overgaard, J. -. Svenning, J. J. Wieringa, B. R. Ramesh, T. Stévant & T. L. P. Couvreur (2016). Remotely sensed temperature and precipitation data improve species distribution modelling in the tropics. *Global Ecology and Biogeography* 25: 443-454.
- Dormann, C. F. (2007). Effects of incorporating spatial autocorrelation into the analysis of species distribution data. *Global Ecology and Biogeography* 16: 129-138.
- Dormann, C. F., M. Bobrowski, D. M. Dehling, D. J. Harris, F. Hartig, H. Lischke, M. D. Moretti, J. Pagel, S. Pinkert, M. Schleuning, S. I. Schmidt, C. S. Sheppard, M. J. Steinbauer, D. Zeuss & C. Kraan (2018). Biotic interactions in species distribution modelling: 10 questions to guide interpretation and avoid false conclusions. *Global Ecology and Biogeography* 27: 1004-1016.
- Elith, J., C. Graham* H., R. Anderson P., M. Dudík, S. Ferrier, A. Guisan, R. Hijmans J., F. Huettmann, J. Leathwick R., A. Lehmann, J. Li, L. Lohmann G., B. Loiselle A., G. Manion, C. Moritz, M. Nakamura, Y. Nakazawa, J. McC. M. Overton, A. Townsend Peterson, S. Phillips J., K. Richardson, R. Scachetti-Pereira, R. Schapire E., J. Soberón, S. Williams, M. Wisz S. & N. Zimmermann E. (2006). Novel methods improve prediction of species' distributions from occurrence data. *Ecography* 29: 129-151.
- Elith, J., J. R. Leathwick & T. Hastie (2008). A working guide to boosted regression trees. *Journal of Animal Ecology* 77: 802-813.
- Elith, J., S. Ferrier, F. Huettmann & J. Leathwick (2005). The evaluation strip: A new and robust method for plotting predicted responses from species distribution models. *Ecological Modelling* 186: 280-289.
- Elith, J. & J. R. Leathwick (2009). Species Distribution Models: Ecological Explanation and Prediction Across Space and Time. *Annual Review of Ecology, Evolution, and Systematics* 40: 677-697.
- Erwin, D. H. (2008). Macroevolution of ecosystem engineering, niche construction and diversity. *Trends in Ecology & Evolution* 23: 304-310.
- Fielding, A. H. & J. F. Bell (1997). A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environmental Conservation* 24: 38-49.

- Fourcade, Y., J. O. Engler, D. Rödder & J. Secondi (2014). Mapping species distributions with MAXENT using a geographically biased sample of presence data: A performance assessment of methods for correcting sampling bias. *PLoS ONE* 9: e97122.
- Franklin, J. (2009). *Mapping Species Distributions: Spatial Inference and Prediction*. 320 s. Cambridge University Press, Cambridge.
- Gallien, L., R. Douzet, S. Pratte, N. E. Zimmermann & W. Thuiller (2012). Invasive species distribution models – how violating the equilibrium assumption can create new insights. *Global Ecology and Biogeography* 21: 1126-1136.
- Gamauf, A., G. Tebb & E. Nemeth (2013). Honey Buzzard *Pernis apivorus* nest-site selection in relation to habitat and the distribution of Goshawks *Accipiter gentilis*. *Ibis* 155: 258-270.
- Gaston, K. J. (2003). *The Structure and Dynamics of Geographic Ranges*. 280 s. OUP Oxford, Oxford.
- Gause, G. F. (1934). *The Struggle for Existence*. Williams and Wilkins, Baltimore.
- Gensbøl, B., P. Koskimies & B. Bertel (1995). *Suomen ja Euroopan päiväpetolinnut*. 408 s. WSOY, Porvoo; Helsinki ; Juva.
- Gilg, O., B. Sittler & I. Hanski (2009). Climate change and cyclic predator–prey population dynamics in the high Arctic. *Global Change Biology* 15: 2634-2652.
- Graham, C. H., S. Ferrier, F. Huettman, C. Moritz & A. T. Peterson (2004). New developments in museum-based informatics and applications in biodiversity analysis. *Trends in Ecology & Evolution* 19: 497-503.
- Guisan, A., C. H. Graham, J. Elith, F. Huettmann, M. Dudik, S. Ferrier, R. Hijmans, A. Lehmann, J. Li, L. G. Lohmann, B. Loiselle, G. Manion, C. Moritz, M. Nakamura, Y. Nakazawa, J. M. Overton, A. T. Peterson, S. J. Phillips, K. Richardson, R. Scachetti-Pereira, R. E. Schapire, S. E. Williams, M. S. Wisz & N. E. Zimmermann (2007). Sensitivity of predictive species distribution models to change in grain size. *Diversity and Distributions* 13: 332-340.
- Guisan, A., T. C. Edwards & T. Hastie (2002). Generalized linear and generalized additive models in studies of species distributions: Setting the scene. *Ecological Modelling* 157: 89-100.
- Guisan, A., B. Petitpierre, O. Broennimann, C. Daehler & C. Kueffer (2014). Unifying niche shift studies: Insights from biological invasions. *Trends in Ecology & Evolution* 29: 260-269.
- Guisan, A. & C. Rahbek (2011). SESAM – a new framework integrating macroecological and species distribution models for predicting spatio-temporal patterns of species assemblages. *Journal of Biogeography* 38: 1433-1444.
- Guisan, A. & W. Thuiller (2005). Predicting species distribution: offering more than simple habitat models. *Ecology Letters* 8: 993-1009.
- Guisan, A., W. Thuiller & N. E. Zimmermann (2017) *Habitat Suitability and Distribution Models: With Applications in R*. 462 s. Cambridge University Press, Cambridge.
- Guisan, A. & N. E. Zimmermann (2000). Predictive habitat distribution models in ecology. *Ecological Modelling* 135: 147-186.
- Hakkarainen, H., S. Mykrä, S. Kurki, R. Tornberg & S. Jungell (2004). Competitive interactions among raptors in boreal forests. *Oecologia* 141: 420-424.
- Hanski, I. (1999). *Metapopulation Ecology*. 313 s. Oxford, UK. Oxford University Press.
- Hao, T., J. Elith, G. Guillera-Arroita & J. Lahoz-Monfort (2019). A review of evidence about use and performance of species distribution modelling ensembles like BIOMOD. *Diversity and Distributions* 25: 839-852.
- He, K. S., B. A. Bradley, A. F. Cord, D. Rocchini, M. -. Tuanmu, S. Schmidtlein, W. Turner, M. Wegmann & N. Pettorelli (2015). Will remote sensing shape the next generation of species distribution models? *Remote Sensing in Ecology and Conservation* 1: 4-18.
- Heikkinen, R. K., M. Luoto & R. Virkkala (2006). Does seasonal fine-tuning of climatic variables improve the performance of bioclimatic envelope models for migratory birds? *Diversity and Distributions* 12: 502-510.
- Heikkinen, R. K., M. Luoto, R. Virkkala, R. G. Pearson & J. -. Körber (2007). Biotic interactions improve prediction of boreal bird distributions at macro-scales. *Global Ecology and Biogeography* 16: 754-763.

- Hernandez, P. A., C. H. Graham, L. L. Master & D. L. Albert (2006). The effect of sample size and species characteristics on performance of different species distribution modeling methods. *Ecography* 29: 773-785.
- Hijmans, R. J. (2012). Cross-validation of species distribution models: removing spatial sorting bias and calibration with a null model. *Ecology* 93: 679-688.
- Hirzel, A. H. & G. Le Lay (2008). Habitat suitability modelling and niche theory. *Journal of Applied Ecology* 45:1372-1381.
- Hirzel, A. H., G. Le Lay, V. Helfer, C. Randin & A. Guisan (2006). Evaluating the ability of habitat suitability models to predict species presences. *Ecological Modelling* 199: 142-152.
- Holt, R. D. & M. Barfield (2009). Trophic interactions and range limits: the diverse roles of predation. *Proceedings of the Royal Society B: Biological Sciences* 276: 1435-1442.
- Holt, R. D., T. H. Keitt, M. A. Lewis, B. A. Maurer & M. L. Taper (2005). Theoretical models of species' borders: single species approaches. *Oikos* 108: 18-27.
- Hutchinson, G. E. (1957). Population Studies. Animal Ecology and Demography: Concluding Remarks. *Cold spring harbor symposia on quantitative biology*. 22: 415-427.
- Ings, T. C., J. M. Montoya, J. Bascompte, N. Blüthgen, L. Brown, C. F. Dormann, F. Edwards, D. Figueroa, U. Jacob, J. I. Jones, R. B. Lauridsen, M. E. Ledger, H. M. Lewis, J. M. Olesen, F. J. F. Van Veen, P. H. Warren & G. Woodward (2009). Review: Ecological networks – beyond food webs. *Journal of Animal Ecology* 78: 253-269.
- Itämes, J. & H. Mikkola (1972). The diet of Honey Buzzards *Pernis apivorus* in Finland. *Ornis Fennica* 49: 7-10.
- Jackson, S. T. & J. T. Overpeck (2000). Responses of plant populations and communities to environmental changes of the late Quaternary. *Paleobiology* 26: 194-220.
- James, G., D. Witten, T. Hastie & R. Tibshirani (2013). *An introduction to statistical learning*. 426 s. Springer, New York.
- Jarnevich, C. S., T. J. Stohlgren, S. Kumar, J. T. Morissette & T. R. Holcombe (2015). Caveats for correlative species distribution modeling. *Ecological Informatics* 29: 6-15.
- Jiménez-Valverde, A. (2012). Insights into the area under the receiver operating characteristic curve (AUC) as a discrimination measure in species distribution modelling. *Global Ecology and Biogeography* 21: 498-507.
- Jiménez-Valverde, A. & J. Lobo (2007). Threshold criteria for conversion of probability of species presence to either or presence-absence. *Acta Oecologica-international Journal of Ecology - ACTA OECOL* 31: 361-369.
- Jiménez-Valverde, A., J. M. Lobo & J. Hortal (2008). Not as good as they seem: the importance of concepts in species distribution modelling. *Diversity and Distributions* 14: 885-890.
- Kadmon, R., O. Farber & A. Danin (2003). A systematic analysis of factors affecting the performance of climatic envelope models. *Ecological Applications* 13: 853-867.
- Kearney, M. (2006). Habitat, environment and niche: what are we modelling? *Oikos* 115: 186-191.
- Kearney, M. & W. Porter (2009). Mechanistic niche modelling: combining physiological and spatial data to predict species' ranges. *Ecology Letters* 12: 334-350.
- Kersalo, J. & Pirinen, P. (2009). Suomen maakuntien ilmasto. Ilmatieteen laitos, Helsinki.
- Kissling, W. D., J. A. Ahumada, A. Bowser, M. Fernandez, N. Fernández, E. A. García, R. P. Guralnick, N. J. B. Isaac, S. Kelling, W. Los, L. McRae, J. Mihoub, M. Obst, M. Santamaria, A. K. Skidmore, K. J. Williams, D. Agosti, D. Amariles, C. Arvanitidis, L. Bastin, F. De Leo, W. Egloff, J. Elith, D. Hobern, D. Martin, H. M. Pereira, G. Pesole, J. Peterseil, H. Saarenmaa, D. Schigel, D. S. Schmeller, N. Segata, E. Turak, P. F. Uhlir, B. Wee & A. R. Hardisty (2018). Building essential biodiversity variables (EBVs) of species distribution and abundance at a global scale. *Biological Reviews* 93: 600-625.
- Kissling, W. D., C. F. Dormann, J. Groeneveld, T. Hickler, I. Kühn, G. J. McNerny, J. M. Montoya, C. Römermann, K. Schiffers, F. M. Schurr, A. Singer, J. Svenning, N. E. Zimmermann & R. B. O'Hara (2012). Towards novel approaches to modelling biotic interactions in multispecies assemblages at large spatial extents. *Journal of Biogeography* 39: 2163-2178.
- Klaassen, R. H. G., M. Hake, R. Strandberg, B. J. Koks, C. Trierweiler, K. -, Exo, F. Bairlein & T. Alerstam (2014). When and where does mortality occur in migratory birds? Direct evidence from long-term satellite tracking of raptors. *Journal of Animal Ecology* 83: 176-184.

- Körner, C. (2007). The use of 'altitude' in ecological research. *Trends in Ecology and Evolution* 22: 569-574.
- Korpela, K., M. Delgado, H. Henttonen, E. Korpimäki, E. Koskela, O. Ovaskainen, H. Pietiäinen, J. Sundell, N. G. Yoccoz & O. Huitu (2013). Nonlinear effects of climate on boreal rodent dynamics: mild winters do not negate high-amplitude cycles. *Global Change Biology* 19: 697-710.
- Kosicki, J. Z. (2017). Should topographic metrics be considered when predicting species density of birds on a large geographical scale? A case of Random Forest approach. *Ecological Modelling* 349: 76-85.
- Kostrzewa, A. (1989). The Effect of Weather on Density and Reproduction Success in Honey Buzzards *Pernis apivorus*. *Raptors in the Modern World*. 187-192.
- Kostrzewa, A. (1998). *Pernis apivorus* Honey Buzzard. 2: 107-120.
- Kostrzewa, A. & R. Kostrzewa (2008). The relationship of spring and summer weather with density and breeding performance of the Buzzard *Buteo buteo*, Goshawk *Accipiter gentilis* and Kestrel *Falco tinnunculus*. *Ibis* 132: 550-559.
- Kremen, C., A. Cameron, A. Moilanen, S. J. Phillips, C. D. Thomas, H. Beentje, J. Dransfield, B. L. Fisher, F. Glaw, T. C. Good, G. J. Harper, R. J. Hijmans, D. C. Lees, E. Louis Jr, R. A. Nussbaum, C. J. Raxworthy, A. Razafimpahanana, G. E. Schatz, M. Vences, D. R. Vieites, P. C. Wright & M. L. Zjhra (2008). Aligning conservation priorities across taxa in Madagascar with high-resolution planning tools. *Science* 320: 222-226.
- Laaksonen, T. & A. Lehikoinen (2013). Population trends in boreal birds: Continuing declines in agricultural, northern, and long-distance migrant species. *Biological Conservation* 168: 99-107.
- Lassau, S. A. & D. F. Hochuli (2007). Associations between wasp communities and forest structure: Do strong local patterns hold across landscapes? *Austral Ecology* 32: 656-662.
- Legendre, P. (1993). Spatial Autocorrelation: Trouble or New Paradigm? *Ecology* 74: 6, 1659-1673.
- Lehikoinen, A., P. Byholm, E. Ranta, P. Saurola, J. Valkama, E. Korpimäki, H. Pietiäinen & H. Henttonen (2009). Reproduction of the common buzzard at its northern range margin under climatic change. *Oikos* 118: 829-836.
- Lehikoinen, A., Jukarainen, A., Mikkola-Roos, M., Below, A., Lehtiniemi, T., Pessa, J., Rajasärkkä, A., Rintala, J., Rusanen, P., Sirkä, P., Tiainen, J., & Valkama, J. (2019). Linnut. 560-570. *Julkaisussa: Hyvärinen, E., Juslén, A., Kemppainen, E., Uddström, A. & Liukko, U.-M. (toim.): Suomen lajien uhanalaisuus – Punainen kirja 2019*. Ympäristöministeriö & Suomen ympäristökeskus. Helsinki.
- Leibold, M. A. (1995). The Niche Concept Revisited: Mechanistic Models and Community Context. *Ecology* 76: 1371-1382.
- Lester, P. J., J. Haywood, M. E. Archer & C. R. Shortall (2017). The long-term population dynamics of common wasps in their native and invaded range. *Journal of Animal Ecology* 86: 337-347.
- Linder, H. P., O. Bykova, J. Dyke, R. S. Etienne, T. Hickler, I. Kühn, G. Marion, R. Ohlemüller, S. J. Schymanski & A. Singer (2012). Biotic modifiers, environmental modulation and species distribution models. *Journal of Biogeography* 39: 2179-2190.
- Lira-Noriega, A. & A. T. Peterson (2014). Range-wide ecological niche comparisons of parasite, hosts and dispersers in a vector-borne plant parasite system. *Journal of Biogeography* 41: 9, 1664-1673.
- Liu, C., G. Newell & M. White (2016). On the selection of thresholds for predicting species occurrence with presence-only data. *Ecology and Evolution* 6: 337-348.
- Lobo, J. M., A. Jiménez-Valverde & R. Real (2008). AUC: a misleading measure of the performance of predictive distribution models. *Global Ecology and Biogeography* 17: 145-151.
- Loiselle, B. A., P. M. Jørgensen, T. Consiglio, I. Jiménez, J. G. Blake, L. G. Lohmann & O. M. Montiel (2008). Predicting species distributions from herbarium collections: does climate bias in collection sampling influence model outcomes? *Journal of Biogeography* 35: 105-116.
- Luoto, M., R. Virkkala & R. K. Heikkinen (2007). The role of land cover in bioclimatic models depends on spatial resolution. *Global Ecology and Biogeography* 16: 34-42.
- Lütolf, M., F. Kienast & A. Guisan (2006). The ghost of past species occurrence: improving species distribution models for presence-only data. *Journal of Applied Ecology* 43: 802-815.
- Mackey, B. G. & D. B. Lindenmayer (2001). Towards a Hierarchical Framework for Modelling the Spatial Distribution of Animals. *Journal of Biogeography* 28: 1147-1166.

- Mattsson, B. J., E. F. Zipkin, B. Gardner, P. J. Blank, J. R. Sauer & J. A. Royle (2013). Explaining Local-Scale Species Distributions: Relative Contributions of Spatial Autocorrelation and Landscape Heterogeneity for an Avian Assemblage. *PLoS ONE* 8: e55097.
- McPherson, J. M., W. Jetz & D. J. Rogers (2004). The effects of species' range sizes on the accuracy of distribution models: ecological phenomenon or statistical artefact? *Journal of Applied Ecology* 41: 811-823.
- McPherson, J. M., W. Jetz & D. J. Rogers (2006). Using coarse-grained occurrence data to predict species distributions at finer spatial resolutions - possibilities and limitations. *Ecological Modelling* 192: 499-522.
- Mearns, R. & I. Newton (1988). Factors Affecting Breeding Success of Peregrines in South Scotland. *Journal of Animal Ecology* 57: 903-916.
- Meier, E. S., F. Kienast, P. B. Pearman, J. Svenning, W. Thuiller, M. B. Araújo, A. Guisan & N. E. Zimmermann (2010). Biotic and abiotic variables show little redundancy in explaining tree species distributions. *Ecography* 33: 1038-1048.
- Merow, C., M. J. Smith, T. C. Edwards Jr, A. Guisan, S. M. McMahon, S. Normand, W. Thuiller, R. O. Wüest, N. E. Zimmermann & J. Elith (2014). What do we gain from simplicity versus complexity in species distribution models? *Ecography* 37: 1267-1281.
- Miller, D. A. W., K. Pacifici, J. S. Sanderlin & B. J. Reich (2019). The recent past and promising future for data integration methods to estimate species' distributions. *Methods in Ecology and Evolution* 10: 22-37.
- Mod, H. K., D. Scherrer, M. Luoto & A. Guisan (2016). What we use is not what we know: Environmental predictors in plant distribution models. *Journal of Vegetation Science* 27: 1308-1322.
- Mueller, A. -, N. Chakarov, H. Hesecker & O. Krüger (2016). Intraguild predation leads to cascading effects on habitat choice, behaviour and reproductive performance. *Journal of Animal Ecology* 85: 774-784.
- Myneni, R. B., C. D. Keeling, C. J. Tucker, G. Asrar & R. R. Nemani (1997). Increased plant growth in the northern high latitudes from 1981 to 1991. *Nature* 386: 698-702.
- Pearce, J. L. & M. S. Boyce (2006). Modelling distribution and abundance with presence-only data. *Journal of Applied Ecology* 43: 405-412.
- Pearman, P. B., A. Guisan, O. Broennimann & C. F. Randin (2008). Niche dynamics in space and time. *Trends in Ecology & Evolution* 23: 149-158.
- Pearson, R. G. & T. P. Dawson (2003). Predicting the impacts of climate change on the distribution of species: are bioclimate envelope models useful? *Global Ecology and Biogeography* 12: 361-371.
- Peterson, A. T., J. Soberón, R. G. Pearson & R. P. Anderson (2011). *Ecological niches and geographic distributions*. 328 s. Princeton University Press, Princeton, NJ.
- Phillips, S. J. & J. Elith (2010). POC plots: calibrating species distribution models with presence-only data. *Ecology* 91: 2476-2484.
- Phillips, S. J., R. P. Anderson & R. E. Schapire (2006). Maximum entropy modeling of species geographic distributions. *Ecological Modelling* 190: 231-259.
- Phillips, S. J., M. Dudík, J. Elith, C. H. Graham, A. Lehmann, J. Leathwick & S. Ferrier (2009). Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data. *Ecological Applications* 19: 181-197.
- Pigot, A. L. & J. A. Tobias (2013). Species interactions constrain geographic range expansion over evolutionary time. *Ecology Letters* 16: 330-338.
- Pokrovsky, I., D. Ehrich, I. Fufachev, R. A. Ims, O. Kulikova, A. Sokolov, N. Sokolova, V. Sokolov & N. G. Yoccoz (2019). Nest association between two predators as a behavioral response to the low density of rodents. *The Auk* 137: 1-13.
- Pollock, L. J., R. Tingley, W. K. Morris, N. Golding, R. B. O'Hara, K. M. Parris, P. A. Vesik & M. A. McCarthy (2014). Understanding co-occurrence by modelling species simultaneously with a Joint Species Distribution Model (JSDM). *Methods in Ecology and Evolution* 5: 397-406.
- Pörtner, H., A. Bennett, F. Bozinovic, A. Clarke, M. Lardies, M. Lucassen, B. Pelster, F. Schiemer & J. Stillman (2006). Trade-Offs in Thermal Adaptation: The Need for a Molecular to Ecological Integration. *Physiological and Biochemical Zoology* 79: 295-313.
- Potapov, E. R. (1997). What Determines the Population Density and Reproductive Success of Rough-Legged Buzzards, *Buteo lagopus*, in the Siberian Tundra? *Oikos* 78: 362-376.

- Pottier, J., Z. Malenovský, A. Psomas, L. Homolová, M. E. Schaepman, P. Choler, W. Thuiller, A. Guisan & N. E. Zimmermann (2014). Modelling plant species distribution in alpine grasslands using airborne imaging spectroscopy. *Biology letters* 10: 20140347.
- Prentice, I. C., W. Cramer, S. P. Harrison, R. Leemans, R. A. Monserud & A. M. Solomon (1992). Special Paper: A Global Biome Model Based on Plant Physiology and Dominance, Soil Properties and Climate. *Journal of Biogeography* 19:117-134.
- Pulliam, H. R. (1988). Sources, Sinks, and Population Regulation. *The American Naturalist* 132: 652-661.
- Pulliam, H. R. (2000). On the relationship between niche and distribution. *Ecology Letters* 3: 349-361.
- Qiao, H., J. Soberón & A. T. Peterson (2015). No silver bullets in correlative ecological niche modelling: insights from testing among many potential algorithms for niche estimation. *Methods in Ecology and Evolution* 6: 10, 1126-1136.
- Randin, C. F., T. Dirnböck, S. Dullinger, N. E. Zimmermann, M. Zappa & A. Guisan (2006). Are niche-based species distribution models transferable in space? *Journal of Biogeography* 33: 1689-1703.
- Reeve, H. K. & P. W. Sherman (1993). Adaptation and the Goals of Evolutionary Research. *The Quarterly review of biology* 68: 1-32.
- Reif, V., R. Tornberg, S. Jungell & E. Korpimäki (2001). Diet variation of common buzzards in Finland supports the alternative prey hypothesis. *Ecography* 24: 267-274.
- Renner, I. W., J. Elith, A. Baddeley, W. Fithian, T. Hastie, S. J. Phillips, G. Popovic & D. I. Warton (2015). Point process models for presence-only analysis. *Methods in Ecology and Evolution* 6: 366-379.
- Ritchie, E. G. & C. N. Johnson (2009). Predator interactions, mesopredator release and biodiversity conservation. *Ecology Letters* 12: 982-998.
- Robinson, B. G., A. Franke & A. E. Derocher (2017). Weather-mediated decline in prey delivery rates causes food-limitation in a top avian predator. *Journal of Avian Biology* 48: 748-758.
- Root, T. (1988). Environmental factors associated with avian distributional boundaries. *Journal of Biogeography* 15: 489-505.
- Saino, N., R. Ambrosini, D. Rubolini, J. Von Hardenberg, A. Provenzale, K. Hüppop, O. Hüppop, A. Lehikoinen, E. Lehikoinen, K. Rainio, M. Romano & L. Sokolov (2011). Climate warming, ecological mismatch at arrival and population decline in migratory birds. *Proceedings of the Royal Society B: Biological Sciences* 278: 835-842.
- Schmidt, N. M., R. A. Ims, T. T. Høye, O. Gilg, L. H. Hansen, J. Hansen, M. Lund, E. Fuglei, M. C. Forchhammer & B. Sittler (2012). Response of an arctic predator guild to collapsing lemming cycles. *Proceedings of the Royal Society B: Biological Sciences* 279: 4417-4422.
- Selås, V. (1997). Nest-site selection by four sympatric forest raptors in southern Norway. *Journal of Raptor Research* 31: 16-25.
- Seyer, Y., G. Gauthier, D. Fauteux & J. -. Therrien (2020). Resource partitioning among avian predators of the Arctic tundra. *Journal of Animal Ecology*.
- Shirley, S. M., Z. Yang, R. A. Hutchinson, J. D. Alexander, K. Mcgarigal & M. G. Betts (2013). Species distribution modelling for the people: Unclassified landsat TM imagery predicts bird occurrence at fine resolutions. *Diversity and Distributions* 19: 855-866.
- Soberón, J. (2007). Grinnellian and Eltonian niches and geographic distributions of species. *Ecology Letters* 10: 12, 1115-1123.
- Soberón, J. & A. Peterson (2005). Interpretation of Models of Fundamental Ecological Niches and Species' Distributional Areas. *Biodiversity Informatics* 2: 1-10.
- Svenning, J. & F. Skov (2004). Limited filling of the potential range in European tree species. *Ecology Letters* 7: 565-573.
- Sylvén, M. (1978). Interspecific Relations between Sympatrically Wintering Common Buzzards *Buteo buteo* and Rough-Legged Buzzards *Buteo lagopus*. *Ornis Scandinavica (Scandinavian Journal of Ornithology)* 9: 197-206.
- Terraube, J., A. Villers, L. Poudré, R. Varjonen & E. Korpimäki (2016). Increased autumn rainfall disrupts predator-prey interactions in fragmented boreal forests. *Global Change Biology* 23: 1361-1373.
- Terraube, J., A. Villers, L. Ruffino, L. Iso-Iivari, H. Henttonen, T. Oksanen & E. Korpimäki (2015). Coping with fast climate change in northern ecosystems: mechanisms underlying the population-level response of a specialist avian predator. *Ecography* 38: 690-699.

- Thibaud, E., B. Petitpierre, O. Broennimann, A. C. Davison & A. Guisan (2014). Measuring the relative effect of factors affecting species distribution model predictions. *Methods in Ecology and Evolution* 5: 947-955.
- Thuiller, W., B. Lafourcade, R. Engler & M. B. Araújo (2009). BIOMOD – a platform for ensemble forecasting of species distributions. *Ecography* 32: 369-373.
- Titeux, N., M. Dufrene, J. Radoux, A. H. Hirzel & P. Defourny (2007). Fitness-related parameters improve presence-only distribution modelling for conservation practice: The case of the red-backed shrike. *Biological Conservation* 138: 207-223.
- Title, P. O. & J. B. Bemmels (2018). ENVIREM: an expanded set of bioclimatic and topographic variables increases flexibility and improves performance of ecological niche modeling. *Ecography* 41: 291-307.
- Tryjanowski, P., T. Pawlikowski, K. Pawlikowski, W. Banaszak-Cibicka & T. H. Sparks (2010). Does climate influence phenological trends in social wasps (Hymenoptera: Vespinae) in Poland? *EJE* 107: 203-208.
- Tylianakis, J. M., R. K. Didham, J. Bascompte & D. A. Wardle (2008). Global change and species interactions in terrestrial ecosystems. *Ecology Letters* 11: 1351-1363.
- Van Horne, B. (1983). Density as a Misleading Indicator of Habitat Quality. *The Journal of Wildlife Management* 47: 893-901.
- Virkkala, R., K. Korhonen, H. Reija & K. Aapala (2000). Metsien ja soiden suojelutilanne metsä- ja suokasvillisuusvyöhykkeittäin valtakunnan metsien 8. inventoinnin perusteella. *Suomen ympäristö* 395: 1-52.
- Virkkala, R., M. Luoto, R. K. Heikkinen & N. Leikola (2005). Distribution patterns of boreal marshland birds: modelling the relationships to land cover and climate. *Journal of Biogeography* 32: 1957-1970.
- Walther, G. -, E. Post, P. Convey, A. Menzel, C. Parmesan, T. J. C. Beebee, J. -, Fromentin, O. Hoegh-Guldberg & F. Bairlein (2002). Ecological responses to recent climate change. *Nature* 416: 389-395.
- Warren, T. L., M. G. Betts, A. W. Diamond & G. J. Forbes (2005). The influence of local habitat and landscape composition on cavity-nesting birds in a forested mosaic. *Forest Ecology and Management* 214: 331-343.
- Waser, N. M., L. Chittka, M. V. Price, N. M. Williams & J. Ollerton (1996). Generalization in Pollination Systems, and Why it Matters. *Ecology* 77: 1043-1060.
- Webster, M. S., P. P. Marra, S. M. Haig, S. Bensch & R. T. Holmes (2002). Links between worlds: Unraveling migratory connectivity. *Trends in Ecology and Evolution* 17: 76-83.
- Whittaker, R. H. (1967). Gradient Analysis of Vegetation. *Biological Reviews* 42: 207-264.
- Wisn, M. S. & A. Guisan (2009). Do pseudo-absence selection strategies influence species distribution models and their predictions? An information-theoretic approach based on simulated data. *BMC ecology* 9: 8.
- Wisn, M., N. Dendoncker, J. Madsen, M. Rounsevell, M. Jespersen, E. Kuijken, W. Courtens, C. Verscheure & F. Cottaar (2008a). Modelling pink-footed goose (*anser brachyrhynchus*) wintering distributions for the year 2050: Potential effects of land-use change in europe. *Diversity and Distributions* 14: 721-731.
- Wisn, M. S., R. Hijmans, J. Li, A. T. Peterson, C. Graham, A. Guisan & NCEAS Predicting Species Distributions Working Group (2008b). Effects of sample size on the performance of species distribution models. *Diversity and Distributions* 14: 763-773.
- Wisn, M. S., J. Pottier, W. D. Kissling, L. Pellissier, J. Lenoir, C. F. Damgaard, C. F. Dormann, M. C. Forchhammer, J. Grytnes, A. Guisan, R. K. Heikkinen, T. T. Høye, I. Kühn, M. Luoto, L. Maiorano, M. Nilsson, S. Normand, E. Öckinger, N. M. Schmidt, M. Termansen, A. Timmermann, D. A. Wardle, P. Aastrup & J. Svenning (2013). The role of biotic interactions in shaping distributions and realised assemblages of species: implications for species distribution modelling. *Biological Reviews* 88: 15-30.
- Wunderlich, R. F., Y. Lin, J. Anthony & J. R. Petway (2019). Two alternative evaluation metrics to replace the true skill statistic in the assessment of species distribution models. *Nature Conservation* 35: 97.
- Zurell, D., J. Elith & B. Schröder (2012). Predicting to new environments: tools for visualizing model behaviour and impacts on mapped distributions. *Diversity and Distributions* 18: 628-634.

Internet-lähteet:

Lintuatlas kertoo pesimälinnuston voittajat ja häviäjät. (2020). Luonnontieteellinen Keskusmuseo, Helsinki. <https://www.luomus.fi/fi/lintuatlas-kertoo-pesimalinnuston-voittajat-haviaajat>.

IUCN (2020). The IUCN Red List of Threatened Species. <https://www.iucnredlist.org>. (viitattu 4.4.2020).

Valkama, Jari, Vepsäläinen, Ville & Lehikoinen, Aleksi 2011: Suomen III Lintuatlas. – Luonnontieteellinen keskusmuseo ja ympäristöministeriö. <http://atlas3.lintuatlas.fi>. Viitattu 15.9.2020.

Aineistot

Lajihavainnot

Suomen Lajitietokeskus/FinBIF (2020). <http://tun.fi/HBF.40368>. Tallennettu 20.4.2020.

Paikkatietoaineistot

Ilmatieteen laitos. Kuukauden keskilämpötila, 1 km x 1 km, 2000-2009. (2020a). CSC – Tieteen tietotekniikan keskus Oy. urn:nbn:fi:csc-kata00001000000000000169. Tallennettu: 16.4.2020.

Ilmatieteen laitos. Kuukauden sademäärä, 1 km x 1 km, 2000-2009. (2020b). CSC – Tieteen tietotekniikan keskus Oy. urn:nbn:fi:csc-kata000010000000000000292. Tallennettu: 16.4.2020.

Ilmatieteen laitos. Kuukauden keskilämpötila, 1 km x 1 km, 2010-2013. (2020c). CSC – Tieteen tietotekniikan keskus Oy. urn:nbn:fi:csc-kata000010000000000000168. Tallennettu. 16.4.2020.

Ilmatieteen laitos. Kuukauden sademäärä, 1 km x 1 km, 2010-2013. (2020d). CSC – Tieteen tietotekniikan keskus Oy. urn:nbn:fi:csc-kata000010000000000000291. Tallennettu: 16.4.2020.

SYKE (osittain METLA,MMM,MML,DVV). CORINE maanpeite 2006, 25m. (2010). https://www.syke.fi/fi-FI/Avoin_tieto/Paikkatietoaineistot/Ladattavat_paikkatietoaineistot. Tallennettu: 16.4.2020.

Ohjelmistot

ArcMap, version 10.7.1. (2019). Environmental Systems Research Institute (ESRI). Redlands, California.

R Core Team. (2016). *R: A Language and Environment for Statistical Computing*, Vienna, Austria. <https://www.R-project.org/>. Viitattu. 25.4.2020.

R-paketit

Greenwell, B., B. Boehmke, J. Cunningham & GBM Developers (2019). gbm: Generalized Boosted Regression Models. R package version 2.1.5.

Broennimann, O., V. Di Cola & A. Guisan (2020). ecospat: Spatial Ecology Miscellaneous Methods. R package version 3.1.

Hijmans, R. J., S. Phillips, J. Leathwick & J. Elith (2017). dismo: Species Distribution Modeling. R package version 1.1-4.

Paradis E., Schliep K (2019). “ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R.” *Bioinformatics*, 35, 526-528.

Thuiller W., D. Georges, R. Engler and F. Breiner (2020). biomod2: Ensemble Platform for Species Distribution Modeling. R package version 3.4.11.

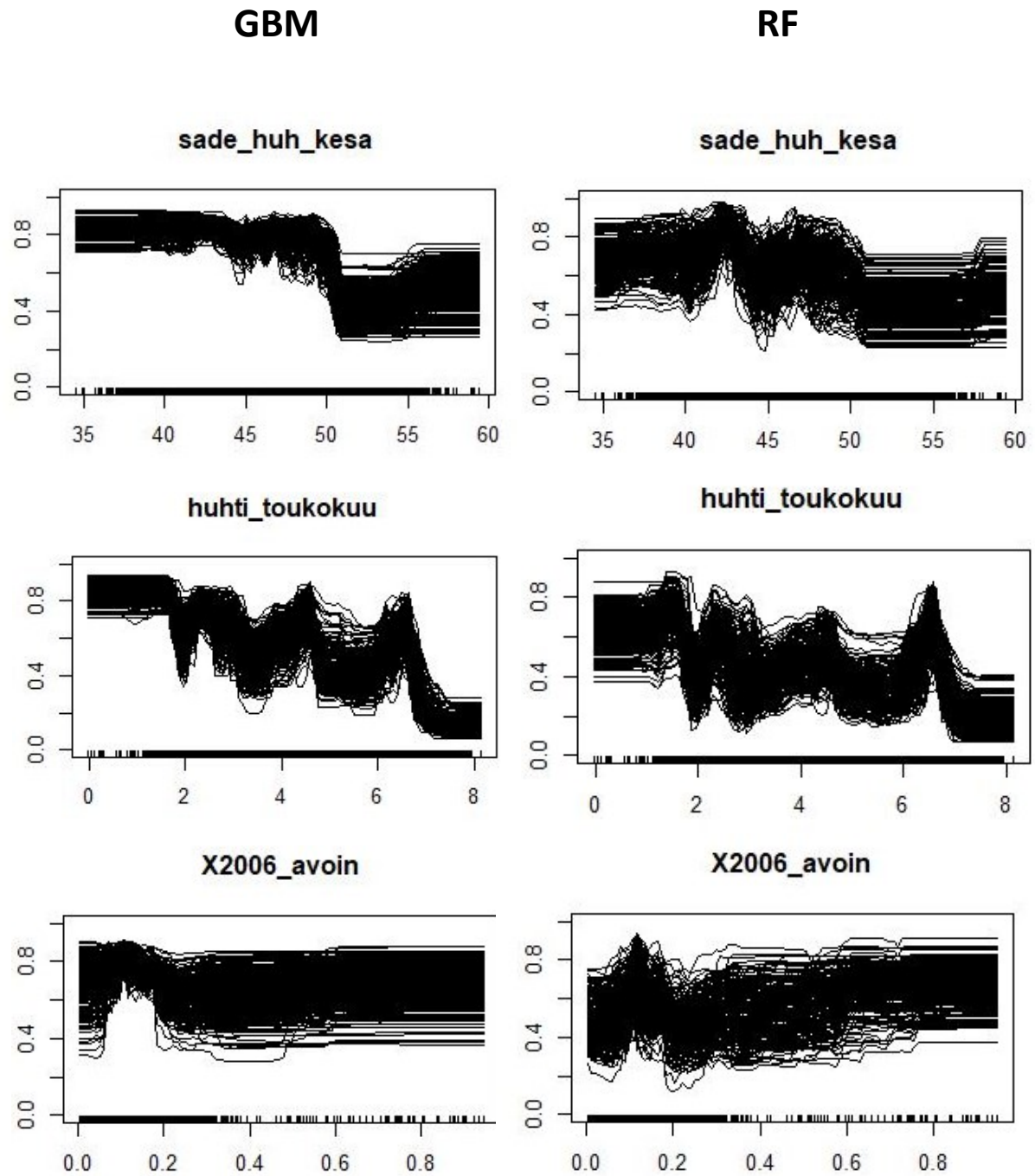
Liitteet

Liite 1. Muuttujien väliset Pearsonin korrelaatiokertoimet (r). Punainen korostus = r yli 0,6.

Suot ja kos- teikot	Metsä	Ihmis- toi- minta	Vesistöt	Maata- lous	Avoin ly.	Sade, huhti- elokuu	Sade, touko- kesäkuu	Sade, kesä- elokuu	Sade, huhti- touko- kuu	Sade, huhti- kesäkuu	Lämpö- tila, huhti- elokuu	Lämpö- tila, touko- kesäkuu	Lämpö- tila, huhti- elokuu	Lämpö- tila, touko- kesä- elokuu	Lämpö- tila, kesä- elokuu
-0,610	0,095	0,512	0,297	0,506	-0,578	0,430	0,393	0,510	-0,107	0,279	0,987	0,987	0,980	0,965	1,000
-0,622	0,130	0,528	0,141	0,593	-0,503	0,323	0,366	0,518	-0,145	0,251	0,985	0,985	0,997	1,000	0,965
-0,621	0,112	0,524	0,228	0,550	-0,549	0,190	0,384	0,518	-0,125	0,269	0,995	0,995	1,000	0,990	0,993
-0,611	0,143	0,520	0,161	0,570	-0,530	0,453	0,395	0,529	-0,111	0,284	0,996	0,996	1,000	0,997	0,980
-0,594	0,158	0,503	0,178	0,534	-0,545	0,399	0,441	0,555	-0,069	0,332	1,000	0,996	0,995	0,985	0,987
-0,104	0,449	-0,081	-0,087	-0,140	-0,197	0,864	0,967	0,780	0,766	1,000	0,332	0,284	0,269	0,251	0,279
0,184	0,397	-0,263	-0,210	-0,278	-0,036	0,912	0,630	0,352	1,000	0,766	-0,069	-0,111	-0,125	-0,145	-0,107
-0,347	0,306	0,089	0,070	0,070	-0,261	0,839	0,855	1,000	0,352	0,780	0,555	0,529	0,518	0,518	0,510
-0,205	0,391	-0,017	-0,006	-0,070	-0,228	0,658	1,000	0,855	0,630	0,967	0,441	0,395	0,384	0,366	0,393
-0,609	0,147	-0,521	-0,014	-0,097	-0,177	1,000	0,658	0,839	0,912	0,864	0,399	0,453	0,190	0,323	0,430
0,236	-0,361	-0,277	-0,318	-0,250	1,000	-0,177	-0,228	-0,261	-0,036	-0,197	-0,545	-0,530	-0,549	-0,503	-0,578
-0,404	-0,222	0,521	-0,133	1,000	-0,250	-0,097	-0,070	0,070	-0,278	-0,140	0,534	0,570	0,550	0,593	0,506
-0,329	-0,430	0,079	1,000	-0,133	-0,318	-0,086	-0,006	0,070	-0,210	-0,087	0,178	0,161	0,228	0,141	0,297
-0,411	-0,224	1,000	0,079	0,521	-0,277	0,521	-0,017	0,089	-0,263	-0,081	0,503	0,520	0,524	0,528	0,512
-0,096	1,000	-0,224	-0,430	-0,222	-0,361	0,147	0,391	0,306	0,397	0,449	0,158	0,143	0,112	0,130	0,095
1,000	-0,096	-0,411	-0,329	-0,404	0,236	-0,609	-0,205	-0,347	0,184	-0,104	-0,594	-0,611	-0,621	-0,622	-0,610

Liite 2. Piekanan vastekäyrät, vaihe 1.

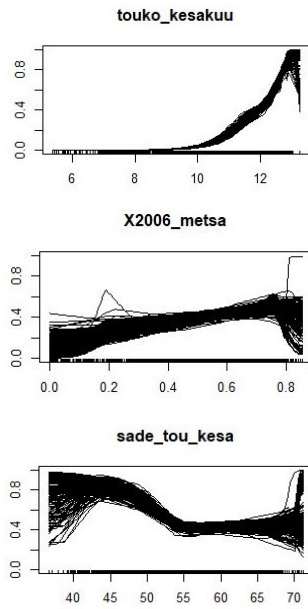
Selite: huhti_toukokuu = huhti-toukokuun lämpötila, sade_huh_kesa = huhti-kesäkuun sademäärä, X2006_avoimen ja harvapuustoisien ympäristön määrä



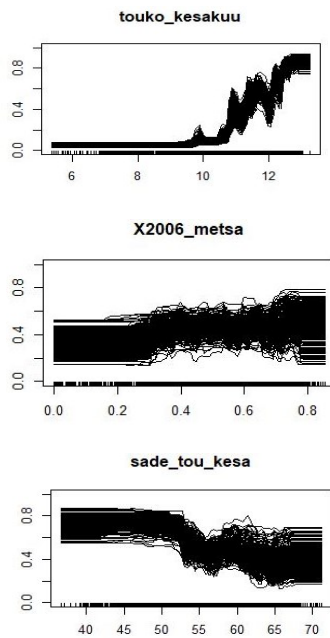
Liite 3. Mehiläishaukan vastekäyrät, vaihe 1.

Selite: touko_kesäkuu= touko-kesäkuun lämpötila, sade_tou_kesa = touko-kesäkuun sademäärä,
X2006_metsa = metsien määrä

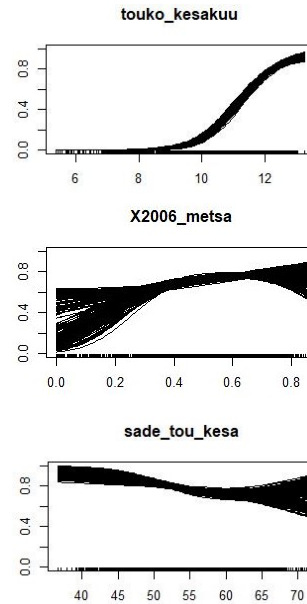
MAXENT



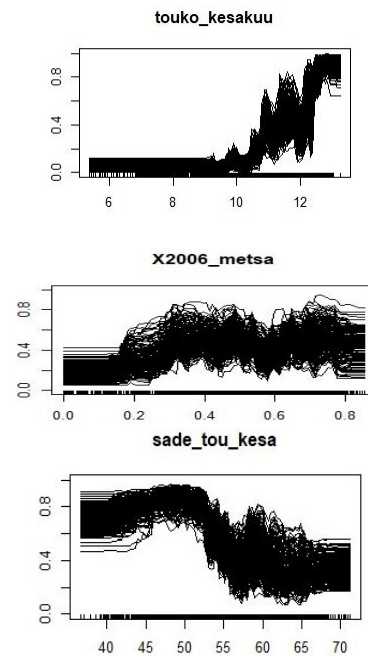
GBM



GLM

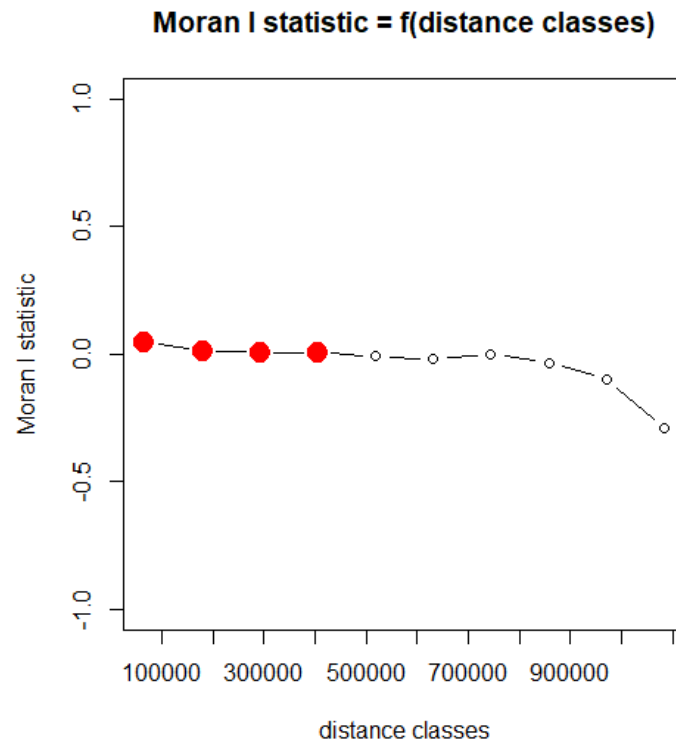


RF

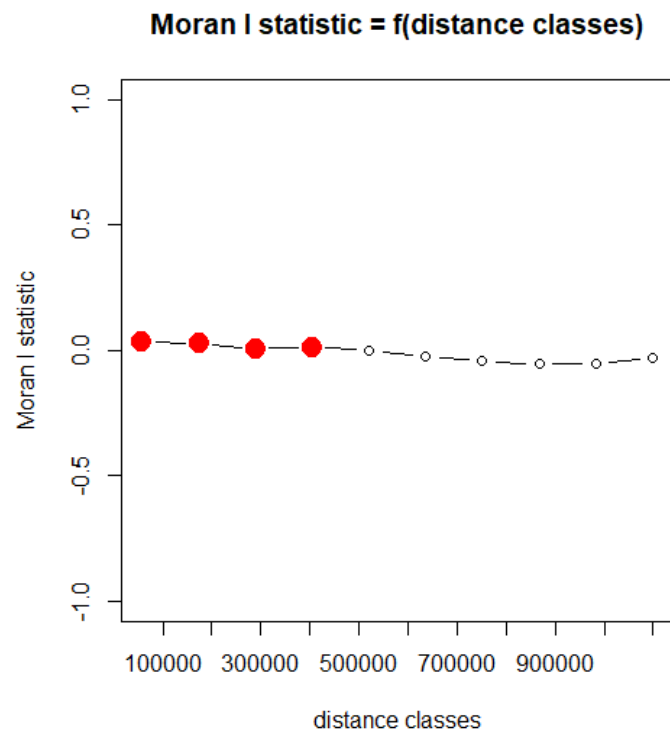


Liite 4. Moranin I kymmenessä etäisyysluokassa, a) piekana ja b) mehiläishaukka. X-akselin arvot metreinä.

a)



b)



Liite 5. Huhti-kesäkuun sademäärä Suomessa.

